



Desmistificando a inteligência artificial

Dora Kaufman

autêntica

VISIT...

LANZAROTE
Caliente.com



Desmistificando a inteligência artificial

Dora Kaufman

autêntica

Desmistificando a inteligência artificial

Dora Kaufman

Desmistificando a inteligência artificial

autêntica

Todos os direitos reservados pela Autêntica Editora Ltda. Nenhuma parte desta publicação poderá ser reproduzida, seja por meios mecânicos, eletrônicos, seja via cópia xerográfica, sem a autorização prévia da Editora.

A autora deste trabalho agradece ao Centro de Inteligência Artificial (C4AI-USP) e o apoio da Fundação de Apoio à Pesquisa do Estado de São Paulo (processo FAPESP #2019/07665-4) e da IBM Corporation.

EDITORAS RESPONSÁVEIS

Rejane Dias

Cecília Martins

REVISÃO

Aline Sobreira

Julia Sousa

CAPA

Diogo Droschi

DIAGRAMAÇÃO

Christiane Moraes de Oliveira

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Kaufman, Dora

Desmistificando a inteligência artificial / Dora Kaufman. -- Belo Horizonte : Autêntica, 2022.

Bibliografia

ISBN 978-65-5928-159-6

1. Automação - Aspectos econômicos 2. Inteligência artificial
3. Inteligência artificial - Inovações tecnológicas 4. Mercado de
trabalho 5. Redes sociais I. Título.

22-101428

CDD-006.3

Índices para catálogo sistemático:

1. Inteligência artificial 006.3

Maria Alice Ferreira - Bibliotecária - CRB-8/7964



Belo Horizonte

Rua Carlos Turner, 420

Silveira . 31140-520

Belo Horizonte . MG

Tel.: (55 31) 3465 4500

São Paulo

Av. Paulista, 2.073 . Conjunto Nacional

Horsa I . Sala 309 . Cerqueira César

01311-940 . São Paulo . SP

Tel.: (55 11) 3034 4468

www.grupoautentica.com.br

SAC: atendimento@grupoautentica.com.br

A expressão “inteligência artificial” foi cunhada há décadas e tem habitado o imaginário humano desde então. Vivemos hoje uma mudança na percepção da sociedade sobre essa tecnologia: notou-se que vários artefatos computacionais podem atingir bom desempenho em tarefas normalmente associadas à inteligência, algo que pareceu por décadas pertencer apenas às obras de ficção científica.

Embora muitas ferramentas computacionais baseadas em inteligência artificial tivessem algum sucesso antes de 2010, foi a partir daquele ano que a tecnologia conseguiu sensibilizar a sociedade como um todo. Por exemplo, antes de 2010 já existiam algoritmos capazes de provar teoremas matemáticos de grande complexidade; porém, a partir de 2010 passamos a ter também sistemas computacionais capazes de identificar faces com grande grau de acerto. Alguns sistemas computacionais contemporâneos têm grau de acerto na detecção de faces *maior* do que o grau de acerto obtido por seres humanos. Ou seja, são sistemas com desempenho “super-humano”. Certamente um provador de teoremas automático gera curiosidade, mas um detector de faces extremamente preciso gera perplexidade. Se um computador pode atingir excelência em uma atividade tão intrinsecamente humana, o que isso nos diz sobre a condição humana?

A década de 2010 foi uma época de crescimento vertiginoso para a inteligência artificial. Em 2010, o economista Kenneth Rogoff, ex-economista-chefe do Fundo Monetário Internacional e professor de Economia em Harvard, escreveu uma coluna no jornal inglês *The Guardian* intitulada “AI Can Power This Decade: As a Former Chess Player I’m Ready to Bet Artificial Intelligence is About to Drive the World’s Economy Forward” (“A IA pode impulsionar esta década: como ex-jogador de xadrez, estou pronto para apostar que a inteligência artificial está prestes a impulsionar a economia mundial”).² Esse título parecia muito otimista até para os pesquisadores da área, pois o histórico da área de inteligência artificial favorecia certa moderação mesmo para os pesquisadores mais confiantes. Com efeito, após grandes promessas feitas por pesquisadores nas décadas de 1950 e 1960, a área de inteligência artificial sofreu um enorme refluxo de financiamento e interesse na década de 1970, sobretudo por ter recebido críticas acerca da disparidade entre promessas e resultados. É famoso o conjunto de predições feitas por Herbert Simon e Allen Newell, dois dos maiores pesquisadores da história da inteligência artificial, em 1957: por exemplo, predições otimistas sobre computadores como campeões de xadrez e compositores

de música em um período de 10 anos.³ Um conhecido relatório produzido em 1972 para o conselho de pesquisa britânico (British Science Research Council), usualmente conhecido como *Lighthill Report*, apresenta a reação a essas promessas, criticando a pesquisa em inteligência artificial e especulando se a área seria viável no longo prazo. Opiniões negativas sobre inteligência artificial se acumularam durante a década de 1970. O entusiasmo retornou durante a década de 1980, com a chegada dos chamados sistemas especialistas, que procuravam reproduzir regras usadas por especialistas em domínios práticos. E o entusiasmo refluíu de novo na década de 1990. Durante aquela década, da qual participei principalmente como estudante de mestrado na Universidade de São Paulo e estudante de doutorado na Universidade Carnegie Mellon (Estados Unidos), muitas novidades apareceram, mas boa parte delas tinha limitado efeito prático. Até mesmo carros autônomos foram construídos, mas não havia ainda a segurança de colocá-los em uso. Havia sistemas de entendimento e geração de fala, de compreensão de imagens, de suporte a decisões e planejamento; a maioria deixava a audiência surpresa e animada, mas não conseguia de fato enfrentar toda a complexidade do mundo real.

Após tantas idas e vindas, a década de 2010 chegou com a combinação certa de poder computacional, acesso a dados, e técnicas de extração automática de padrões em dados. Um momento particularmente importante foi a vitória de redes neurais artificiais “profundas” em uma competição de sistemas automáticos de interpretação de imagens, em 2012. Redes neurais artificiais procuram reproduzir computacionalmente alguns aspectos do sistema nervoso humano, combinando unidades de processamento simples (os “neurônios artificiais”) em camadas que se ligam de forma inspirada nas sinapses do cérebro humano.

A capacidade de redes neurais artificiais de resolver problemas práticos foi muito questionada durante os anos 2000, quando alternativas mais eficientes e precisas foram desenvolvidas. A partir de 2012 isso mudou, com o uso de redes artificiais com muitas camadas – daí o adjetivo “profundas”. A tecnologia de redes neurais artificiais profundas avançou tanto que hoje muitos confundem o seu uso, muitas vezes rotulado de “aprendizado profundo” (*deep learning*), com toda a área de inteligência artificial. O aprendizado profundo representa uma das principais partes da subárea preocupada com aprendizado de máquina, em que se procura ensinar o computador a realizar tarefas a partir de dados, de instruções, de textos.

Mas note que a busca por inteligência artificial depende de avanços não só na capacidade de extrair padrões de grandes massas de dados, mas também na capacidade de receber instruções complexas no tempo. Nós, humanos, não aprendemos apenas observando grandes massas de dados. Nós lemos livros e manuais; nós vamos à escola para assistir a aulas e absorver os conhecimentos dos mais experientes.

E, além de aprendizado de máquina, técnicas de planejamento automático e

de representação de conhecimento são fundamentais na construção de inteligências artificiais. Enquanto décadas passadas viram grande ênfase em métodos baseados na representação de conhecimento a partir da venerável teoria de lógica, que remonta a Aristóteles, hoje se procura combinar afirmações, argumentos, fatos e probabilidades de forma flexível e pragmática.

O progresso tem sido surpreendente em todos esses tópicos, embora computadores ainda estejam distantes da habilidade humana de argumentar, de negociar, de interagir com interlocutores, de... uma grande lista de atividades. Para quem trabalha na área de inteligência artificial, é evidente que estamos ainda muito longe dos cenários de ficção científica – tanto os cenários utópicos quanto os distópicos.

Diante dessa ebulição tecnológica, a sociedade se debate entre otimismo e pessimismos. Por um lado, a inteligência artificial pode mudar nossa vida de forma substancial, aumentando nossa produtividade e oferecendo suporte e auxílio sempre que necessário. O efeito de alguns avanços é claro: por exemplo, poderemos ter cadeias logísticas mais eficientes ao usarmos técnicas de planejamento automático; com isso, poderemos economizar energia e ter mais produtos à disposição de compradores interessados. Outros avanços são mais sutis: poderemos ter música ambiente personalizada, criada artificialmente, atendendo a nossos humores em todos os instantes. Quanto vale esse conforto?

A sociedade também se vê diante de variados receios. Alguns ainda pertencem à ficção científica: estarão os robôs contra nós em um futuro distante? Mas outros receios são concretos. Poderá um sistema computacional autônomo tomar decisões que afetem a vida de seres humanos? Que regras devem ser seguidas por agentes artificiais? Em caso de falha, quem é o culpado? Além disso, como controlar um sistema computacional para evitar invasões de privacidade? E como garantir que o ser humano mantenha o controle democrático sobre sua sociedade, não perdendo esse controle para entidades artificiais de difícil acesso? Como lidar com a perda de postos de trabalho à medida que algumas tarefas sejam substituídas via automação? E como tudo isso afeta nossa identidade humana?

Todas essas questões merecem debate. Por um lado, os medos precisam ser entendidos e dissipados, quando for o caso, ou discutidos e remediados. Por outro lado, o potencial da tecnologia precisa ser entendido e explorado em seus aspectos positivos. A tecnologia de inteligência artificial não é perfeita, como toda tecnologia; com um debate sério, podemos minimizar seus problemas e maximizar seus benefícios.

Em resumo, é fundamental *desmistificar a inteligência artificial*. Por isso, este livro merece ser lido com cuidado. Seus capítulos viajam por temas interdisciplinares, examinando a definição de inteligência artificial e seus componentes, discutindo o mercado de trabalho, as questões éticas, os aspectos econômicos e regulatórios, bem como as aplicações e os impactos em áreas como

saúde e clima. Escrito em linguagem acessível, o texto não se furta a questionar tanto os conceitos básicos de inteligência artificial quanto as percepções da sociedade sobre essa tecnologia.

A autora e colega, Dora Kaufman, tem trabalhado há tempos na interface entre tecnologia e sociedade. Temos tido a chance de interagir no âmbito do Centro de Inteligência Artificial (Center for Artificial Intelligence – c4ai), criado em 2020 com suporte da ibm e da Fundação de Amparo à Pesquisa do Estado de São Paulo (fapesp), sediado na Universidade de São Paulo (usp) e com a Pontifícia Universidade Católica de São Paulo (puc-sp) como instituição parceira. Um dos objetivos do centro é justamente oferecer um espaço de debate amplo e interdisciplinar, onde diferentes ideias possam ser exploradas e testadas. A profa. Dora exercita muito bem esse elemento de interdisciplinaridade que é essencial para o entendimento das inteligências artificiais que estão entre nós e das que estarão no futuro. Parabéns a ela pelo texto lícido; parabéns ao leitor pela escolha de leitura.

Introdução

Em palestra proferida em 1985, Richard Feynman, Prêmio Nobel de 1965 e um dos mais reconhecidos físicos teóricos, debate temas críticos do campo da inteligência artificial.⁴ O diálogo com o público tem início com uma pergunta-chave: “Você acha que haverá uma máquina que possa pensar como os humanos e ser mais inteligente do que os humanos?”. Para Feynman, as futuras máquinas não pensarão como os seres humanos, da mesma forma que um avião não voa como os pássaros. Entre outras diferenças, os aviões não batem asas; são processos, dispositivos e materiais distintos. Quanto à questão de as máquinas superarem a inteligência humana, na visão do físico, o ponto de partida está na própria definição de “inteligência”.

De fato, é difícil definir o que entendemos por “inteligência”. Segundo Stuart Russell, pesquisador que é referência no campo da inteligência artificial, uma entidade é inteligente na medida em que o que faz é capaz de alcançar o que deseja.⁵ Russell escreve: “Todas essas outras características da inteligência – perceber, pensar, aprender, inventar e assim por diante – podem ser compreendidas por meio de suas contribuições para nossa capacidade de agir com sucesso”. Russell lembra que o conceito de inteligência, desde os primórdios da filosofia grega antiga, está associado a capacidades humanas (perceber, raciocinar e agir), o que não seria o caso da inteligência artificial, “meros” modelos de otimização com objetivos definidos pelos humanos, e não dotados desses atributos. Outros autores não consideram a “inteligência” uma prerrogativa humana, como o próprio Marvin Minsky, um dos fundadores do campo da ia, ao argumentar que os sistemas de inteligência artificial têm habilidades, apesar de limitadas, de aprendizagem e raciocínio. Complicando ainda mais esse debate, as técnicas atuais de ia lidam com percepção, análise de texto, processamento de linguagem natural (pnl), raciocínio lógico, sistemas de apoio à decisão, análise de dados e análise preditiva.

Outro tema abordado por Feynman em sua palestra de 1985 foi o reconhecimento de padrões em grandes conjuntos de dados, à época um desafio ainda não totalmente viabilizado por técnicas empíricas de inteligência artificial. A programação computacional, pondera o físico, não contemplaria as nuances da realidade, como luminosidades, distâncias e ângulos de inclinação da cabeça num conjunto de fotos – os seres humanos são capazes de reconhecer uma pessoa pelo

movimento do corpo ao andar, pela maneira como mexe no cabelo e por outros pequenos e sutis detalhes.

Resolver tarefas executadas pelos humanos intuitivamente, e com relativo grau de subjetividade, era um desafio dos primórdios do campo da inteligência artificial. Várias tentativas que envolviam linguagens formais apoiadas em regras de inferência lógica tiveram êxito limitado, sugerindo a necessidade de os sistemas gerarem seu próprio conhecimento pela extração de padrões de dados, ou seja, “aprender” com os dados sem receber instruções explícitas. Esse processo é usualmente denominado de “aprendizado de máquina” (*machine learning*), subcampo da inteligência artificial criado em 1959 e hoje certamente o maior da ia em número de praticantes.

A técnica de aprendizado de máquina que melhor resolve esses desafios é o aprendizado profundo (*deep learning*), que introduz representações complexas, expressas em termos de outras representações mais simples organizadas em diversas camadas. Essa estrutura codifica uma função matemática que mapeia conjuntos de valores de entrada (*inputs*) para valores de saída (*outputs*); redes com maior profundidade (mais camadas) têm apresentado resultados positivos em várias áreas, particularmente em visão computacional e reconhecimento de voz e imagem.

Essa relativamente nova técnica de aprendizado de máquina é denominada de “redes neurais de aprendizado profundo” (*deep learning neural networks*, dl_{nn}) pela inspiração no funcionamento do cérebro biológico. A técnica é capaz de lidar com dados de alta dimensionalidade, por exemplo, milhões de *pixels* num processo de reconhecimento de imagem. Além disso, seus algoritmos estabelecem correlações nos dados não perceptíveis aos desenvolvedores humanos, origem do problema da interpretabilidade ou “caixa-preta”.

Na última década, a disponibilidade de grandes conjuntos de dados (*big data*), produzidos por uma sociedade hiperconectada, e a maior capacidade computacional, particularmente com o advento das *gpus* (*graphic processing units*), geraram resultados positivos nos sistemas baseados nessa técnica. Assim, a técnica de aprendizado profundo tornou-se fator estratégico de processos decisórios pela capacidade de gerar *insights* preditivos com taxas relativamente altas de acurácia, permeando a maior parte das aplicações atuais de inteligência artificial. Entretanto, a técnica ainda possui limitações, como requerer grandes quantidades de dados de qualidade para desenvolvimento, treinamento e aperfeiçoamento dos modelos, e demandar, por sua arquitetura complexa, *hardware* com grande capacidade de processamento (intensivo em consumo de energia, consequentemente, em emissão de CO_2).

Na esfera ética, destaca-se o problema do viés nos resultados (ou resultados discriminatórios por gênero, raça, etnia, entre outros). Em geral, atribui-se o viés às bases de dados tendenciosas, porém, o viés pode emergir antes da coleta de dados, em função das decisões tomadas pelos desenvolvedores (as variáveis

contempladas no modelo, inclusive, determinam a seleção dos dados).

Em verificações estatísticas de rotina nos dados de um grande hospital, por exemplo, pesquisadores da Universidade da Califórnia, em Berkeley, coordenados pelo médico e professor Ziad Obermeyer, especializado na interseção entre aprendizado de máquina e saúde, constataram que, em um sistema automatizado de triagem para assistência médica de alta complexidade, pacientes autoidentificados como negros tendiam a receber pontuações de risco mais baixas do que pacientes brancos em condições similares. Amplamente utilizado em hospitais e seguradoras nos Estados Unidos, o sistema invariavelmente discriminava os pacientes negros: apenas 17,7% dos pacientes selecionados eram negros, quando a proporção não tendenciosa seria de 46,5%.

Após um minucioso escrutínio, os pesquisadores detectaram na variável-chave inicial, definida pelos desenvolvedores do sistema, a origem do viés: correlação entre custo anual com saúde e urgência no atendimento, ou seja, quanto maiores os gastos em saúde no ano anterior, maior a probabilidade de o paciente precisar de cuidados médicos de alta complexidade. Dois argumentos indicam o erro na escolha dessa variável: a) em condições de saúde similares, um paciente branco norte-americano custa em média 1.800 dólares a mais por ano do que um paciente negro, contudo, essa diferença deve-se ao acesso desfavorável dos pacientes negros aos serviços de saúde; e b) um paciente pode ter um histórico médico saudável no ano anterior (logo, custo zero) e necessitar de tratamento urgente e de alta complexidade no momento da consulta.

A descoberta foi relatada à Optum, criadora do sistema, que repetiu a análise, encontrando os mesmos resultados; na sequência, em colaboração, as duas equipes introduziram novas variáveis, reduzindo em 84% o viés original. Os resultados do estudo foram publicados em outubro de 2019 na prestigiosa revista *Science*.⁶

Nesse caso, frequentemente citado no debate sobre “discriminação algorítmica”, o viés foi claramente associado a decisões humanas, e agravado pelo racismo sistêmico (as equipes médicas e os gestores dos hospitais, predominantemente brancos, não se deram conta da evidente discriminação). No estágio atual de desenvolvimento da inteligência artificial, a subjetividade humana está presente na criação dos sistemas, no treinamento dos algoritmos, na escolha da base de dados, na verificação e nos ajustes, e na visualização e interpretação dos resultados.

Distorções desse tipo, comuns nos sistemas de ia, remetem à importância da diversidade na formação das equipes desenvolvedoras desses sistemas, agregando, inclusive, experiências e conhecimentos de fora do campo tecnológico. Pelas próprias características (e formação) dos desenvolvedores, o foco dos projetos tecnológicos está na funcionalidade dos sistemas, visando solucionar problemas práticos, e em geral não contemplam os impactos éticos e sociais.

Marvin Minsky, em 1970, ao ganhar o Prêmio Turing (conhecido como o

Nobel da computação), previu que dentro de “três a oito anos teremos uma máquina com a inteligência geral de um ser humano médio”, profecia que está longe de se concretizar. Contudo, a inteligência artificial está mediando a vida cotidiana dos cidadãos do século xxi; gradativamente, os algoritmos de ia estão substituindo os humanos na execução de inúmeras tarefas, fortemente presentes em sistemas de decisão automatizados.⁷

Nesse sentido, é mandatório que os usuários intermediários – profissionais de saúde, profissionais de educação, gestores de rh, gestores financeiros, e muitos outros – adquiram noções básicas da lógica e do funcionamento da inteligência artificial para, inclusive, capacitar-se a fazer as perguntas críticas aos fornecedores de tecnologia. Ao contratar ou adotar um sistema de ia, dispor de conhecimento é essencial para identificar e evitar resultados tendenciosos.

O primeiro desafio é enfrentar as deficiências em nossa formação. Herdamos um sistema de ensino baseado na lógica da economia industrial, compartimentalizado e especializado, que forma profissionais de ciências exatas pouco sensíveis à ética e ao social, e profissionais de ciências sociais e humanas resistentes à tecnologia. A compartimentalização dificulta, inclusive, a formação de equipes interdisciplinares. A experiência mostra que não basta juntar profissionais de várias áreas, é necessário construir pontes para superar potenciais conflitos de linguagem, de raciocínio, de metodologia de análise, de objetivos e prioridades.

São imensas as barreiras dos profissionais que lidam com a conexão entre a tecnologia e o domínio de aplicação, os “habitantes das terras fronteiriças”, por exemplo, os “conectores” entre os desenvolvedores dos sistemas de inteligência artificial para a saúde e os profissionais de saúde. Os pacientes, legitimamente, demandam dos médicos justificativas para os procedimentos previstos (recomendados) pelos algoritmos de ia, tornando as fronteiras um lugar complicado no qual seus residentes precisam se capacitar para traduzir linguagens e visões de mundo díspares.

Scott Hartley, no livro *O fuzzy e o techie: por que as ciências humanas vão dominar o mundo digital*, argumenta a favor das ciências humanas num mundo dominado pela tecnologia, defendendo a parceria entre as ciências exatas e as humanas, em que as primeiras focam no “como fazer” da revolução tecnológica, e as segundas, no “por quê”, “para quê” e “quando”.⁸ Hartley observa que os fundadores e dirigentes de empresas de tecnologia do Vale do Silício, em geral, não têm formação básica em tecnologia e matriculam seus filhos em escolas humanistas que enfatizam a curiosidade intelectual, a criatividade, a comunicação interpessoal, a empatia e a capacidade de aprendizagem e resolução de problemas, ou seja, as chamadas *soft skills*, cada vez mais valorizadas no mercado de trabalho. A formação em humanidades tem se mostrado essencial para liderar a inovação, de produtos a modelos de negócios.

Os algoritmos de inteligência artificial são bons em identificar padrões

estatísticos, mas eles não têm como saber o que esses padrões significam, porque estão confinados ao “*math world*” (mundo da matemática). Sem compreender o mundo real, a ia não tem como avaliar se os padrões estatísticos que encontram são coincidências úteis ou sem sentido. Como alertam alguns especialistas, o perigo real hoje não é que a inteligência artificial seja mais inteligente do que os humanos, mas supor que ela seja mais inteligente do que os humanos e, conseqüentemente, confiar nela para tomar decisões importantes. A inteligência artificial atual deveria ser meramente um parceiro dos especialistas humanos.

Não é o caso de todos nos tornarmos especialistas em ia, mas apenas de adquirirmos uma familiaridade básica que permita aperfeiçoar nossa interação com a tecnologia, maximizando os benefícios e minimizando os potenciais danos. Numa analogia com os automóveis, em geral os motoristas desconhecem como fabricar um carro ou mesmo como consertá-lo, no entanto, sabem que frente a um obstáculo – um semáforo, outro carro, um pedestre – é preciso frear, e não acelerar, para evitar acidente. Temos de saber quando “frear” e quando “acelerar” a inteligência artificial.

Em 2016, quando decidi me dedicar aos impactos éticos e sociais da inteligência artificial, tracei como meta estudar, simultaneamente, a própria tecnologia, decisão que se mostrou acertada e que recomendo fortemente. A primeira iniciativa foi procurar meu amigo de infância Davi Geiger, cientista em ia radicado nos Estados Unidos há mais de 40 anos – PhD no Instituto de Tecnologia de Massachusetts (mit) e atualmente professor titular no Courant Institute of Mathematical Sciences, da Universidade de Nova York –, que se tornou meu mentor na tecnologia.

A cada encontro com o Davi, suas explicações são traduzidas em desenhos esquemáticos. Adquiri o hábito de guardar esses desenhos para depois retomá-los e constatar minha evolução no entendimento da inteligência artificial. Essa interação requer de mim um grande esforço, particularmente porque a linguagem da tecnologia é matemática, num grau de sofisticação do qual meus estudos matemáticos anteriores não dão conta. Quando os desenhos começam a fazer sentido para mim, o desafio subsequente é traduzi-los em linguagem natural, palatável ao campo das ciências sociais e humanas e, posteriormente, aos interessados em geral.

Em outubro do mesmo ano, por recomendação do próprio Davi, participei do meu primeiro evento internacional de inteligência artificial, a conferência Ethics of Artificial Intelligence,⁹ organizada por David Chalmers e Ned Block, filósofos da Universidade de Nova York. Foram dois dias de debates intensos, com a participação de 300 pesquisadores eminentes, palestrantes ou público debatedor, num espectro amplo que ia desde o filósofo inglês Nick Bostrom, autor do livro *Superintelligence*, que abriu a conferência, até Yann LeCun, à época chefe de inteligência artificial do Facebook, e o Prêmio Nobel Daniel Kahneman (com quem tive a honra de tomar um café no Village, dois dias depois). A

conferência originou o livro *Ethics of Artificial Intelligence* (2020), coletânea de 17 ensaios inéditos organizada pelo filósofo S. Matthew Liao, sobre o qual fiz uma resenha para a revista *teccogs*, do Programa Tecnologias da Inteligência e Design Digital (tidd), da puc-sp.¹⁰ Sem dúvida, foi uma excepcional boas-vindas ao campo da inteligência artificial.

Os resultados positivos dos modelos empíricos de ia ainda são recentes, e o mercado editorial tem relativamente poucos títulos sobre isso, particularmente no Brasil, e menos ainda em linguagem acessível ao público leigo. Este livro pretende preencher essa lacuna e transmitir os fundamentos e a lógica da inteligência artificial por meio da associação a temáticas correntes. Trata-se de uma coletânea de artigos da minha coluna na revista *Época Negócios*, publicados entre junho de 2019 e abril de 2022, agrupados em 12 grandes temas: “Fundamentos e lógica da ia”, “ia e mercado de trabalho”, “Justiça e ética na ia”, “A ia e o problema do viés”, “A Economia de Dados e o poder das *big techs*”, “Iniciativas regulatórias”, “ia e saúde”, “ia no combate à covid-19”, “A ia e o clima”, “ia e cultura”, “Interação humano-máquina” e “Pensando o futuro”. Cada parte é precedida de uma pequena apresentação.

As tecnologias não são todas iguais, algumas adicionam valor incremental à sociedade e outras são disruptivas. Ao reconfigurar a lógica de funcionamento da economia e aportar inéditos modelos de negócios, as disruptivas provocam períodos de reorganização no que Joseph Schumpeter denominou de “destruição criativa”. As tecnologias de propósito geral (*general purpose technologies*, gpt) estão nesse último bloco. São tecnologias-chave, moldam toda uma era e reorientam as inovações nos setores de aplicação, como a máquina a vapor, a eletricidade e o computador. A inteligência artificial é a tecnologia de propósito geral do século xxi e tende a impactar cada vez mais a vida em sociedade. Com a ia migramos de um mundo dominado por máquinas programadas para um mundo de máquinas probabilísticas, implicando lógicas e riscos distintos. O mundo da inteligência artificial é bem mais complexo, temos de aprender a habitar esse mundo para continuar sendo relevantes, profissional e socialmente.

Fundamentos e lógica da IA

Algoritmos de ia estão em toda parte. Numa sociedade hiperconectada, vivemos em ambientes técnico-sociais inteligentes em que a sociabilidade e a comunicação geram dados digitais. A ia já domina o mercado de ações, compõe música, produz arte, dirige carros, escreve artigos de notícias, prognostica tratamentos médicos, decide sobre crédito e contratação, recomenda entretenimento, e tudo isso ainda em seus primórdios.

Na última década, a ia tornou-se a tecnologia de propósito geral do século xxi. A tendência é que a lógica da ia torne-se hegemônica na geração de riqueza, criando um valor econômico sem precedentes. Estamos migrando, aceleradamente, para a Economia de Dados, ou Capitalismo de Dados, ou Capitalismo “Dadocêntrico”, termos que expressam um modelo econômico cuja matéria-prima estratégica são os dados.

Os filmes de ficção científica, e as narrativas subsequentes, confundem a fronteira entre ficção e realidade. A ia hoje é “mera” função matemática que, ao ser capaz de lidar com o *big data*, gradativamente, vem assumindo o protagonismo nas relações socioeconômicas. A subjetividade humana, contudo, é decisiva em todas as etapas do desenvolvimento e da interpretação dos resultados. São os especialistas humanos que constroem os modelos, definem os parâmetros, criam as bases de dados, selecionam o domínio da aplicação; os preconceitos e valores humanos estão presentes em cada decisão de cada etapa do processo.

Como toda tecnologia, a ia é social e humana, seus efeitos dependem do que os seres humanos fazem com ela, como a percebem, como a experimentam e usam, como a inserem nos ambientes técnico-sociais. Cabe à sociedade humana deliberar, dentre inúmeras questões, sobre se a ia deve ser aplicada em todos os domínios e para executar todas as tarefas, e se o uso da ia em aplicações de alto risco se justifica. O desafio é buscar o equilíbrio entre mitigar (ou eliminar) os riscos e preservar o ambiente de inovação, sem supervalorizar nem demonizar a ia.

Este bloco contempla sete artigos que introduzem os fundamentos básicos da ia, desde as distinções entre redes neurais e cérebro humano até elementos críticos como as limitações intrínsecas à técnica e a relevância da qualidade da base de dados.

No filme *Matrix* (Lilly e Lana Wachowski, 1999), o *hacker* Neo descobre que o mundo é mera simulação de computador, e que as máquinas inteligentes declararam guerra à humanidade. No filme britânico *Ex-Machina* (Alex Garland, 2015), um funcionário do buscador Bluebook se defronta com robôs com inteligência artificial e tem com eles diálogos assustadores. Em palestra em São Paulo, o filósofo francês Edgar Morin, com a lucidez de seus 97 anos, enfatizou que o papel das artes é de sensibilizar e fazer compreender. No caso da inteligência artificial, contudo, a ficção confunde mais do que elucida: não há nenhum indício científico de um futuro com robôs subjugando os humanos.

A ia faz parte da nossa vida cotidiana. Acessamos sistemas inteligentes para programar o itinerário com o Waze, pesquisar no Google e receber da Netflix e do Spotify recomendações de filmes e músicas. A Amazon captura nossas preferências no fluxo de dados que coleta a partir das nossas interações com a plataforma. A Siri, da Apple, e a Alexa, da Amazon, são assistentes pessoais digitais inteligentes que nos ajudam a localizar informações úteis com acesso por meio de voz.

Os algoritmos de inteligência artificial mediam as interações nas redes sociais, como a seleção do que será publicado no *feed* de notícias do Facebook. Eles estão igualmente presentes nos diagnósticos médicos, nos sistemas de vigilância, na prevenção a fraudes, nas análises de crédito, nas contratações de rh, na gestão de investimento, na indústria 4.0, no atendimento automatizado (*chatbot*); bem como nas estratégias de marketing, nas pesquisas, na tradução de idiomas, no jornalismo automatizado, nos carros autônomos, no comércio físico e virtual, nos canteiros de obras, nas perfurações de petróleo, na previsão de epidemias. Estamos na era da personalização, viabilizada pela extração das informações contidas nos dados que geramos em nossas movimentações *online*.

A maioria dos avanços observados na última década provém do modelo chamado de *deep learning* (aprendizado profundo), técnica de *machine learning* (aprendizado de máquina), subárea da inteligência artificial, que consiste em técnicas estatísticas que permitem que as máquinas “aprendam” com os dados (e não sejam programadas).

Na década de 1980, um grupo restrito de pesquisadores – Geoffrey Hinton, Yoshua Bengio e Yann LeCun –, inspirado no funcionamento do cérebro biológico, propôs o caminho das redes neurais para o aprendizado de máquina, obtendo reconhecimento definitivo em 2012, ao vencer a competição ImageNet.

Nos anos seguintes, elas se tornaram onipresentes, recebendo expressivos investimentos das gigantes de tecnologia que incorporaram a inteligência artificial em seus modelos de negócio. Cerca de 100 serviços do Google, por exemplo, usam ia.

Deep learning é um modelo estatístico de previsão de cenários futuros e a probabilidade de eles se realizarem e quando; a denominação provém da profundidade das camadas que formam a arquitetura das redes neurais. Correlacionando grandes quantidades de dados, os algoritmos de ia são capazes de estimar com mais assertividade a probabilidade de um tumor ser de um determinado tipo de câncer, ou a probabilidade de uma imagem ser de um cachorro, ou a previsão de quando um equipamento necessitará de reposição, ou o candidato apropriado para determinada função, ou o tipo de serviço ou produto adequado aos desejos do consumidor.

No estágio atual da ia, não se trata de ensinar as máquinas a pensar, mas apenas a prever a probabilidade de os eventos ocorrerem, por meio de modelos estatísticos e grandes quantidades de dados. Esses sistemas carecem da essência da inteligência humana: capacidade de compreender o significado. Apesar de todos os esforços, houve pouco progresso em prover a ia de senso intuitivo, de capacidade de formar conceitos abstratos e de fazer analogias e generalizações.

Conceitos básicos sobre o mundo – tridimensionalidade, movimentação e permanência dos objetos, gravidade, inércia e rigidez – são aprendidos pelos seres humanos essencialmente pela observação. Descobrir como incorporar esse aprendizado às máquinas é a chave para o progresso da inteligência artificial, ou seja, as máquinas serem capazes de aprender como o mundo funciona assistindo a um vídeo do YouTube.

A tecnologia avança aceleradamente, mas ainda distante da condição humana

31.10.2019

Em Hamburgo, importante cidade portuária no norte da Alemanha, em 2017, a polícia foi chamada pelos vizinhos porque a Alexa, assistente virtual da Amazon, “estava dando uma festa com o som altíssimo”. O dono do apartamento, Oliver Haberstroh, que na ocasião estava bebendo cerveja num bar, ao voltar para casa encontrou uma nova fechadura; na delegacia do bairro lhe entregaram as novas chaves e uma fatura. A Amazon, após minuciosa investigação, alegou que a Alexa foi ativada remotamente, e o volume aumentou através do aplicativo de *streaming* de música móvel de terceiros, mas mesmo assim ofereceu pagar o custo do incidente. As duas hipóteses – defeito do assistente virtual ou ativação remota – não configuram autonomia, livre-arbítrio, agenciamento, nenhum dos atributos que caracterizam os seres humanos.

O Atlas é um humanoide criado pela Boston Dynamics em 2013, capaz de reproduzir movimentos humanos, tais como saltar, girar no ar, dar cambalhotas, todos efeitos do campo da robótica. Trata-se de um sistema de controle avançado que, usando algoritmos de otimização, automatizou alguns desses movimentos. A Sophia, criada em 2016 pela Hanson Robotics, reconhecida como a fabricante de robôs “mais humanos”, é dotada de expressividade, estética, interatividade, pele maleável, e tem dezenas de computadores acoplados que permitem simular uma gama completa de expressões faciais, rastrear o rosto do interlocutor, reconhecer faces e imitar as expressões faciais de outras pessoas. O Atlas simula o sistema motor humano, a Sophia simula o sistema cognitivo humano; o aprendizado é o elemento comum entre os dois sistemas: construídos para aprender como aprender.

As máquinas inteligentes estão sendo concebidas para, com base em grandes conjuntos de dados, “aprender” por meio de processos não totalmente explicáveis, isto é, os desenvolvedores dessas máquinas não sabem exatamente como elas aprendem para desempenhar as tarefas (a chamada “caixa-preta”, que não deve ser confundida com autonomia). O cientista da computação Davi Geiger alerta que aqui reside, talvez, a maior questão ética na inteligência artificial, o fato de não sabermos o que e como as máquinas realmente aprendem, não deixando de lembrar que também não sabemos o que e como exatamente os humanos aprendem.

Os modelos de inteligência artificial amplamente utilizados são chamados de “redes neurais”, porque são inspirados no funcionamento do cérebro biológico. Simplificadamente, no cérebro ocorrem continuamente impulsos

nervosos (sinais químicos e elétricos) que são conduzidos até o próximo neurônio, num fenômeno conhecido como sinapse, que transmite a informação entre as camadas de neurônios (com mais precisão: a sinapse refere-se à interrupção que ocorre entre as duas camadas de neurônios). Cada neurônio tem uma espécie de antena, chamada de dendrito, que é o canal de entrada da informação. Os modelos de redes neurais reproduzem essa lógica (não são concebidos para executar tarefas a partir de equações predefinidas, a programação tradicional), e o nome *deep learning* (aprendizado profundo) vem do fato de que possuem várias camadas de processamento compostas de neurônios artificiais (cada camada aperfeiçoa uma parte da informação). Segundo o neurocientista Roberto Lent, “mesmo as alternativas oferecidas pela inteligência artificial, que podem propiciar retornos ‘inteligentes’ de reciprocidade aos aprendizes, não atingiram ainda a riqueza de possibilidades das interações entre humanos”.¹¹

O ponto a ressaltar é que mesmo os sistemas mais sofisticados, como o da Sophia, não chegam nem a tangenciar a complexidade do funcionamento do cérebro biológico. O aprendizado humano depende de um grande número de neurônios que, por sua vez, formam circuitos complexos responsáveis pela nossa estrutura cognitiva e comportamental. Roberto Lent, em seu livro *O cérebro aprendiz*, apresenta-nos alguns números ilustrativos dessa complexidade: cada ser humano tem 86 bilhões de neurônios, e cada neurônio recebe cerca de 100 mil sinapses (ordem de grandeza total na casa do quatrilhão); a partir da décima semana de gestação, a produção de novos neurônios no cérebro humano em desenvolvimento atinge aproximadamente a velocidade de 250 mil novas células por minuto; as tecnologias atuais não permitem estudar o cérebro humano em nível microscópico, levando os cientistas a recorrerem a animais: o tempo de computação e análise utilizado por pesquisadores chineses para estudar as conexões de 135 mil neurônios de uma mosca foi de 10 anos, estimando-se em 17 milhões de anos o tempo necessário para o mesmo procedimento no cérebro humano.¹²

A inteligência artificial hoje é fundamentalmente modelos estatísticos que, baseados em dados, calculam a probabilidade de eventos ocorrerem. Esse pequeno avanço tem sido responsável por transformações na economia, nas relações pessoais, na sociedade em geral, mas estamos a léguas de distância da chamada *general AI* (ou *strong AI* ou *full AI*), que, supostamente, seria uma inteligência artificial dotada de capacidades de nível humano.

Máquinas “pensantes” e autônomas de seus criadores povoam a literatura e a ciência há muito tempo. Em 1818, a escritora inglesa Mary Shelley publicou *Frankenstein ou o Prometeu moderno*, considerado a primeira obra de ficção científica. O romance relata a história do estudante de ciências naturais Victor Frankenstein, que, no empenho em descobrir os mistérios da criação, dedica-se a conceber um ser humano gigantesco. O monstro foge do laboratório e refugia-se numa floresta, onde aprende a sobreviver. Frankenstein tem inteligência própria, autonomia e até mesmo sentimentos, semelhantemente aos robôs dos filmes de ficção científica.

Para citar outro romance, em 1872, o escritor britânico Samuel Butler publicou *Erewhon*, considerado a primeira especulação sobre a possibilidade de máquinas conscientes.¹³ Na fictícia cidade de Erewhon não existem máquinas, fruto da percepção amplamente compartilhada pelos erewonianos de que elas são potencialmente perigosas. Num conflito antigo entre os cidadãos pró-máquinas e os antimáquinas, os últimos foram vitoriosos: “Veja, essas máquinas estão ficando cada vez mais sofisticadas e capazes e nos lançarão de surpresa no cativeiro. Nós nos tornaremos subservientes às máquinas e, eventualmente, seremos descartados”.

O debate permeia igualmente a ciência, abordado por vários cientistas e historiadores nos últimos dois séculos. O matemático inglês Alan Turing, no texto de 1951 “Can Digital Computers Think?” (“Os computadores digitais podem pensar?”), pondera sobre os argumentos racionais a favor e contra a ideia de que as máquinas possam ser levadas a pensar, e sobre a analogia entre máquina e cérebro.¹⁴ Segundo ele, admitir que ambos sejam análogos envolveria assumir também que as máquinas sejam dotadas de livre-arbítrio, o que não faz sentido no caso de um computador digital programado (ele mesmo, contudo, questiona esse atributo como diferencial, aventando a hipótese de que o sentimento de livre-arbítrio dos seres humanos seja mera ilusão). A inteligência não deve ser confundida com a aleatoriedade, potencialmente responsável pela capacidade das máquinas de nos surpreender ao entregarem resultados distintos da programação (ou distintos da intenção inicial).

Avançando no tempo, o historiador Yuval Harari, em seu *Homo Deus: uma breve história do amanhã*, alinhado com a visão de que “consciência” é um atributo humano, advoga que o advento das máquinas inteligentes representa um descolamento entre inteligência e consciência, gerando dois tipos de inteligência:

a inteligência consciente e a inteligência não consciente.¹⁵ Com essa restrição, Harari impõe um limite ao progresso da ia: as máquinas inteligentes, ao não serem dotadas de consciência, nunca vão competir com a inteligência humana.

O avanço da inteligência artificial intensificou o debate sobre quais são os limites das máquinas, tanto na ficção quanto na ciência. Deixando de lado o futuro da ia, mantendo o foco nos modelos estatísticos de probabilidade (modelos de *machine learning* e *deep learning*), que são a maior parte das aplicações atuais de ia, Stuart Russell, um dos mais renomados cientistas da computação, em seu livro *Human Compatible: Artificial Intelligence and the Problem of Control* (na edição brasileira, 2021, *Inteligência Artificial a Nosso Favor: como manter o controle sobre a tecnologia*),¹⁶ propõe pensar sobre o que significa “perder o controle” sobre as máquinas.

Para Russell, a crença de que estaríamos na eventualidade de perder o controle sobre as “máquinas inteligentes” baseia-se em um erro inicial na definição de ia. Para começar, as máquinas não são “inteligentes” no sentido dado pelos seres humanos – ser capaz de agir para atingir objetivos próprios –, pelo contrário, elas não têm objetivos, são os seres humanos que imputam seus objetivos nos sistemas inteligentes (são máquinas de otimização). O que não impede, no entanto, que as máquinas inteligentes encontrem soluções melhores do que os seres humanos, o que é uma realidade com as tecnologias de ia.

Russell propõe abandonar a ideia de “máquinas inteligentes” em favor de “máquinas benéficas”, na medida em que se espera que suas ações atinjam os objetivos dos seres humanos, melhorando a vida em sociedade. “O que temos hoje é uma relação binária, uma propriedade do sistema composto pela máquina e pelos seres humanos, com resultados superiores aos obtidos individualmente. As máquinas funcionam como uma espécie de metade de um sistema combinado com os seres humanos”, argumenta.

A autodenominação “*Homo sapiens*” (homem sábio) expressa a crença dos seres humanos de que sua superioridade esteja no fato de serem os únicos seres vivos dotados de inteligência. Não apenas satisfeitos com essa aparente “dádiva divina”, buscamos entender como funciona o cérebro humano, o que é consciência, quais os mecanismos produtores do pensamento e, mais ainda, estamos empenhados em dotar as máquinas de ao menos parte desses atributos. Se algum dia as máquinas serão efetivamente inteligentes, por enquanto, pertence ao campo da ficção.

Redes neurais artificiais e a complexidade do cérebro humano

14.8.2020

A ideia de usar a lógica de aprendizagem em uma máquina remete, ao menos, a Alan Turing. Em seu artigo seminal de 1950 “Computing Machinery and Intelligence” (“Maquinaria computacional e inteligência”), em que propõe um teste para a pergunta se uma máquina pode pensar, Turing cogita a ideia de produzir um programa que, em vez de simular a mente do adulto, simule a mente de uma criança. Evoluindo ao longo do tempo, ele a chamou de “máquina-criança”.¹⁷

O campo da inteligência artificial foi inaugurado num seminário de verão, em 1956, com a premissa de que “todos os aspectos da aprendizagem ou qualquer outra característica da inteligência podem, em princípio, ser descritos tão precisamente que uma máquina pode ser construída para simulá-la”.¹⁸ Quase 70 anos depois, a ia ainda está restrita a modelos empíricos, o campo não possui uma teoria, e é controversa a associação de conceitos como inteligência e aprendizado a máquinas.

Apostando na superação das limitações científicas atuais, um grupo de líderes do Vale do Silício está empenhado em “vencer a morte”, atingir o que eles chamam de “amortalidade”. Ray Kurzweil, no livro *A singularidade está próxima*, prevê que ao final do século xxi a parte não biológica da inteligência humana será trilhões de vezes mais poderosa que a inteligência humana biológica, e não haverá distinção entre humanos e máquinas.¹⁹ Em 2013, o Google fundou a Calico, empresa dedicada a “resolver a morte”, em seguida nomeou Bill Maris, igualmente empenhado na busca da imortalidade, como presidente do fundo de investimento Google Venture, que aloca 36% do total de 2 bilhões de dólares em *startups* de biociência com projetos associados a prorrogar a vida. No mesmo ano, Peter Diamandis, cofundador e presidente executivo da Singularity University, lançou a empresa Human Longevity, dedicada a combater o envelhecimento, projetando que o aumento da longevidade criaria um mercado global de 3,5 trilhões de dólares.

A *startup* Neuralink, fundada por Elon Musk em 2016, investe no desenvolvimento de uma interface cérebro-computador que possibilitaria, por exemplo, fazer *streaming* de música diretamente no cérebro; outro foco é viabilizar a transferência da mente humana para um computador, libertando o cérebro do corpo envelhecido e acoplando-o a uma “vida digital”, num processo chamado “*mind-upload*” (transferência da mente humana). Essa visão utópica pós-humanista supõe que esses melhoramentos conduzirão à vitória sobre o

envelhecimento biológico, portanto, ao nascimento de uma nova espécie: os pós-humanos, libertados de seu corpo mortal.

Na visão de Yoshua Bengio, um dos três idealizadores da técnica de IA *deep learning*: “Esses tipos de cenários não são compatíveis com a forma como construímos atualmente a ia. As coisas podem ser diferentes em algumas décadas, não tenho ideia, mas, no que me diz respeito, isso é ficção científica”.²⁰

O ponto de partida para avaliar quão distante a ciência está dessas ideias é compreender a arquitetura e o funcionamento do cérebro. O neurocientista Roberto Lent, em recente conversa no ciclo de debates do tiddigital, traduziu a extrema complexidade do cérebro humano em números: cada ser humano possui 86 bilhões de neurônios e 85 bilhões de células coadjuvantes no processo da informação.²¹ Considerando apenas os neurônios, como em média ocorrem 100 mil sinapses por neurônio, temos um total aproximado de 8,6 quadrilhões de circuitos que, ainda por cima, são plásticos, ou seja, mutáveis continuamente.

Numa sinapse, transmissão da informação de um neurônio para outro, o segundo neurônio pode bloquear a informação, modificar a informação, aumentar a informação, ou seja, a informação que passa para o segundo neurônio pode ser bastante diferente daquela que entrou, indicando a enorme capacidade do cérebro em modificar a informação. As regiões responsáveis pela memória e pelas emoções, entre outros fatores, afetam a informação inicial.

O aprendizado de uma criança, que alguns comparam com o aprendizado de máquina, ocorre por complexos processos cerebrais. Segundo Lent, para aprender a escrever, uma criança precisa formar uma conexão entre escrita e significado, e para isso usa a região do hemisfério esquerdo do cérebro, antes usada para o reconhecimento de faces. Isso acontece porque não temos uma área cerebral da escrita e da leitura; essas habilidades são construtos da civilização que têm “apenas” quatro mil anos, logo não houve tempo evolutivo suficiente para haver uma área cerebral específica. A região de reconhecimento de faces, desenvolvida nos primeiros anos do bebê, desloca-se, portanto, do hemisfério esquerdo para o direito, e ali, no hemisfério esquerdo, começa a ser implantada uma região de reconhecimento de símbolos da escrita. Isso mostra o grau de plasticidade do cérebro, ao realocar funções que vão aparecendo durante a vida do indivíduo com novas aquisições culturais.

A neuroplasticidade – capacidade do cérebro de mudar, adaptar-se e moldar-se em nível estrutural e funcional, quando sujeito a novas experiências do ambiente interno e externo – gera uma complexidade que é difícil reproduzir em uma máquina. A dinâmica do cérebro é altamente modulável, não é uma cadeia de informação linear que leva diretamente a um resultado previsível, como também nos ensinou Roberto Lent.

Para Yann LeCun, outro dos três idealizadores da técnica de *deep learning*, a observação e a interação da criança com o mundo desempenham um papel central no aprendizado infantil, incluindo a noção de que o mundo é

tridimensional, tem gravidade, inércia e rigidez. Esse tipo de acúmulo de enorme quantidade de conhecimento é que não se sabe como reproduzir nas máquinas – observar o mundo e descobrir como ele funciona.

Andrew Ng, respeitado cientista e empreendedor em inteligência artificial, crê que o maior problema da ia seja de comunicação: “O tremendo progresso por meio da ia ‘estreita’ está fazendo com que as pessoas argumentem erroneamente que há um tremendo progresso na agi (*artificial general intelligence*). Francamente, não vejo muito progresso em direção à agi”.²² A inteligência artificial que atualmente permeia aplicativos, plataformas *online*, sistemas de rastreamento e de reconhecimento facial, diagnósticos médicos, modelos de negócio, redes sociais, plataformas de busca, otimização de processos, *chatbots* e mais uma infinidade de tarefas automatizáveis é apenas um modelo estatístico de probabilidade baseado em dados, “anos-luz” distante da complexidade do cérebro humano.

Da personalização do discurso em Aristóteles à personalização com algoritmos de inteligência artificial

11.9.2020

Os algoritmos de inteligência artificial atuam como curadores da informação, personalizando, por exemplo, as respostas nas plataformas de busca, como o Google, e a seleção do que será publicado no *feed* de notícias de cada usuário do Facebook. O ativista Eli Pariser reconhece a utilidade de sistemas de relevância, ao fornecer conteúdo personalizado, mas alerta para os efeitos negativos da formação de “bolhas”, ao reduzir a exposição a opiniões divergentes.²³ Para Cass Sunstein, esses sistemas são responsáveis pelo aumento da polarização cultural e política, pondo em risco a democracia.²⁴ Existem muitas críticas a esses sistemas, algumas justas, outras nem tanto. O fato é que personalização, curadoria, clusterização, mecanismos de persuasão, nada disso é novo, cabe é investigar o que mudou com a ia.

A personalização do discurso, por exemplo, remete a Aristóteles. A arte de conhecer o ouvinte e adaptar o discurso ao seu perfil, não para convencê-lo racionalmente, mas para conquistá-lo pelo “coração”, é o tema da obra *Retórica*. Composta de três volumes, o Livro ii é dedicado ao plano emocional, listando as emoções que devem conter um discurso persuasivo: ira, calma, amizade, inimizade, temor, confiança, vergonha, desvergonha, amabilidade, piedade, indignação, inveja e emulação. Para o filósofo, todos, de algum modo, praticam a retórica na sustentação de seus argumentos. Essa obra funda as bases da retórica ocidental, que, com seus mecanismos de persuasão, busca influenciar o interlocutor, seja ele usuário, consumidor, cliente ou eleitor.

Cada modelo econômico tem seus próprios mecanismos de persuasão, que extrapolam motivações comerciais com impactos culturais e comportamentais. Na economia industrial, caracterizada pela produção e pelo consumo massivo de bens e serviços, a propaganda predominou como meio de convencimento nas decisões dos consumidores, inicialmente tratados como uma massa de indivíduos indistinguíveis. O advento das tecnologias digitais viabilizou a comunicação segmentada em função de características, perfis e preferências similares, mas ainda distante da hipersegmentação proporcionada pelas tecnologias de inteligência artificial.

A hipersegmentação com algoritmos de ia é baseada na mineração de grandes conjuntos de dados (*big data*) e sofisticadas técnicas de análise e previsão, particularmente os modelos estatísticos de redes neurais/*deep learning*. Esses modelos permitem extrair dos dados informações sobre seus usuários e consumidores e fazer previsões com relativamente alto grau de acurácia – desejos,

comportamentos, interesses, padrões de pesquisa, por onde circulam, bem como a capacidade de pagamento e até o estado de saúde. Os algoritmos de ia transformam em informação útil a imensidão de dados gerados nas movimentações *online*.

Na visão de Shoshana Zuboff, a maior ameaça não está nos dados produzidos voluntariamente em nossas interações nos meios digitais (“dados consentidos”), mas nos “dados residuais”, sob os quais os usuários de plataformas *online* não exercem controle.²⁵ Até 2006, os dados residuais eram desprezados, mas, com a sofisticação dos modelos preditivos de inteligência artificial, tornaram-se valiosos: a velocidade de digitalização, os erros gramaticais cometidos, o formato dos textos, as cores preferidas e mais uma infinidade de detalhes do comportamento dos usuários são registrados e inseridos na extensa base de dados, gerando projeções sobre o comportamento humano atual e futuro. Outro aspecto ressaltado por Zuboff é que as plataformas tecnológicas, em geral, captam mais dados do que o necessário para a dinâmica de seus modelos de negócio, ou seja, para melhorar produtos e serviços, e os utilizam para prever o comportamento de grupos específicos (“excedente comportamental”).

Esses processos de persuasão ocorrem em níveis invisíveis, sem conhecimento ou consentimento dos usuários, que desconhecem o potencial e a abrangência das previsões dos algoritmos de inteligência artificial; num nível mais avançado, essas previsões envolvem personalidade, emoções, orientação sexual e política, ou seja, um conjunto de informações que em tese não era a intenção do usuário revelar. As fotos postadas nas redes sociais, por exemplo, geram os chamados “sinais de previsão”, tais como os músculos e a simetria da face, informações utilizadas no treinamento de algoritmos de ia de reconhecimento de imagem.

A escala atual de geração, armazenamento e mineração de dados, associada aos modelos assertivos de personalização, é um dos elementos-chave da mudança de natureza dos atuais mecanismos de persuasão. Comparando os modelos tradicionais com os de algoritmos de inteligência artificial, é possível detectar a extensão dessa mudança: de mensagens elaboradas com base em conhecimento superficial e limitado do público-alvo, a partir do entendimento das características generalistas das categorias, para mensagens elaboradas com base no conhecimento profundo e detalhado, minucioso, do público-alvo, hipersegmentação e personalização; de correlações entre variáveis determinadas pelo desenvolvedor do sistema para correlações entre variáveis determinadas automaticamente com base nos dados dos usuários; de recursos limitados para associar comportamentos *offline* e *online* para a capacidade de capturar e armazenar dados de comportamento *offline* e agregá-los aos dados capturados *online*, formando uma base de dados única, mais completa, mais diversificada, mais precisa; de mecanismos de persuasão visíveis (propaganda na mídia) e relativamente visíveis (propaganda na internet) para mecanismos de persuasão

invisíveis; de baixo grau de assertividade para alto grau de assertividade; de instrumentos limitados de medição/verificação dos resultados para instrumentos mais precisos; de limitada capacidade preditiva de tendências futuras para capacidade com grau de acurácia média em torno de 80-90%, com possibilidade de previsão de quando essas tendências podem se realizar; e de reduzida capacidade de distorcer imagem e voz para enorme capacidade de distorcer imagem e voz, com as *deep fakes*.

Como sempre, cabe à sociedade encontrar um ponto de equilíbrio entre os benefícios e as ameaças da inteligência artificial. No caso, entre a proteção aos direitos humanos civilizatórios e a inovação e o avanço tecnológico, e entre a curadoria da informação e a manipulação do consumo, do acesso à informação e dos processos democráticos.

Interpretabilidade e confiança: variáveis interdependentes nos modelos de redes neurais

9.10.2020

No campo da inteligência artificial, a técnica de aprendizado de máquina conhecida como redes neurais de aprendizado profundo (*deep learning neural networks*, dl_{nn}) está sendo cada vez mais usada na tomada de decisões médicas, particularmente quando envolvem reconhecimento de imagem, uma das áreas mais bem-sucedidas de implementação da ia. A lógica é semelhante ao processo de investigação do médico: correlaciona os sintomas descritos pelo paciente com o “banco de dados” armazenado em sua memória, acrescido de pesquisas em plataformas médicas, buscando semelhanças que levem ao diagnóstico. Os algoritmos de ia, no caso de reconhecimento de imagem de uma tomografia, por exemplo, tendem a identificar anomalias com mais assertividade porque a correlacionam com uma base maior de dados e captam sinais imperceptíveis a outros processos.

A técnica de redes neurais, contudo, já possui intrinsecamente uma variável de incerteza por ser um modelo estatístico de probabilidade. Os cientistas estão empenhados em reduzir a chamada caixa-preta (*black-box*), ou seja, serem capazes de explicar como o sistema chegou a determinada previsão e, preferencialmente, de forma que o usuário compreenda e possa até identificar os erros cometidos pelos algoritmos. Existe um *trade-off* entre precisão e interpretabilidade: quanto maior a precisão, menor a transparência em relação ao seu funcionamento. Diferentemente das sugestões da Netflix ou do Waze, em que uma recomendação equivocada não tem maiores consequências, na medicina a opacidade da técnica contribui, legitimamente, para a resistência dos profissionais de saúde em adotá-la.

A arquitetura dessa técnica é composta de várias camadas intermediárias (camadas “escondidas”, daí advém o nome de redes neurais profundas), que interpretam detalhes de uma imagem não perceptíveis aos seres humanos (padrões invisíveis). A alta dimensionalidade dos modelos (valores e quantidades de *pixels*, por exemplo, no reconhecimento de imagem) requer uma matemática complexa, agravando ainda mais a dificuldade de compreensão pelos usuários (na verdade, transcende a capacidade da cognição humana).

Os dados também afetam o grau de confiabilidade nos resultados. Como o sistema estabelece correlações com base nos dados, a fonte e a qualidade deles é fator crítico (além das questões éticas, como privacidade). Se o usuário suspeitar das fontes de dados, dificilmente levará em conta em suas decisões as previsões do sistema de inteligência artificial, particularmente, importante enfatizar, em áreas

sensíveis como saúde.

Qualquer máquina é concebida para funcionar e, antes de ser comercializada, é submetida a uma série de testes, em que são apurados indicadores confiáveis, com percentuais de acerto em níveis aceitáveis para considerá-la aprovada. No caso de máquinas de inteligência artificial, o processo é bem mais complexo, por conta da opacidade (caixa-preta). A chamada interpretabilidade do sistema de ia é a tentativa de entender e determinar qual grau de confiança atribuir ao resultado obtido. Por exemplo, quando se classificam “cachorro” e “gato”, temos duas unidades de saída da rede, uma representando um cachorro, e a outra, um gato. Dado um *input*, digamos, uma imagem de um gato, a unidade que representa gato *light up* (fica com um valor perto de 1), enquanto a unidade que representa cachorro não *light up* (fica com um valor perto de zero). Assim, a rede conclui que é gato. Quanto mais a unidade do gato é perto de 1, e quanto mais a unidade de cachorro é perto de 0, mais se tem confiança de que se trata de um gato.

Existem métodos mais sofisticados; por exemplo, pode-se olhar para unidades dentro da rede, não no *output* (saída/resultado), permitindo visualizar que unidades são mais ativadas quando o sistema funciona bem nos testes e que unidades são mais ativadas quando não funciona bem nos testes, e assim pode-se “entender melhor” se vai funcionar ou não para um dado de entrada. Por exemplo: se fotos de animais sem orelha são *input* no sistema, e nota-se que certas unidades que sempre são *light up* não são mais *light up* quando não há orelha, aí pode-se interpretar essas unidades como detectoras da existência de orelha.

Estão em curso esforços científicos para criar interpretações amigáveis do funcionamento dos modelos de redes neurais, ou seja, tornar acessíveis sistemas complexos. Um dos caminhos que vem sendo investigado é por meio de exemplos; o propósito é ajudar o usuário a entender o resultado e determinar o grau de confiabilidade em um sistema “imperfeito”, contribuindo para inferir porque um algoritmo gerou um determinado *output*. É o que recomendam as diretrizes do *Guidebook*, do Google,²⁶ aos seus desenvolvedores e designers com o propósito de melhorar a confiabilidade de seus produtos para os usuários.

O guia defende que sejam formuladas e explicitadas explicações parciais sobre o funcionamento do modelo, mesmo que deixem de contemplar as partes mais complexas, de difícil entendimento para o usuário médio. A aposta da People + ai Research (pair), área responsável pelo guia, coliderada pela brasileira Fernanda Viégas, é que quanto maior for a transparência, maior será o potencial de influenciar positivamente a experiência do usuário, aumentando a chance de ele tomar uma decisão com base na recomendação do modelo.

Essa técnica de inteligência artificial amplamente usada oferece diversos benefícios, mas tem limitações. O mais prudente é que seus usuários não confiem plenamente nos seus resultados. Em primeiro lugar, porque são técnicas estatísticas de probabilidade, logo, possuem grau de incerteza intrínseco, e, em

segundo, por conta da opacidade de seu funcionamento (como confiar plenamente em algo que não se domina, não se compreende?), além das várias limitações técnicas. A inteligência artificial implementada atualmente em larga escala deve ser encarada como parceira dos profissionais humanos nos processos de decisão, e não soberana, ou seja, capaz de contribuir para aumentar a inteligência humana especializada, e não substituí-la.

Base de dados para treinar algoritmo de IA não é salsicha: qualidade dos “ingredientes” é crítica

12.11.2021

“O mundo ao nosso redor é cada vez mais coreografado pelos algoritmos de inteligência artificial”, sentencia o jornalista de tecnologia da *The New Yorker* Matthew Hutson, em artigo publicado na prestigiada revista *ieee Spectrum*.²⁷ A ia está no cerne dos modelos de negócio das plataformas e aplicativos tecnológicos e das decisões automatizadas, logo mediando a vida dos cidadãos do século xxi. Pela aparente assertividade de seus resultados, a inteligência artificial tem sido aplicada indiscriminadamente, sem avaliação e controle de riscos. Sistemas de ia não auditados são utilizados em áreas sensíveis, com impacto direto na vida das pessoas.

A técnica de aprendizado de máquina (*machine learning*), que permeia a maior parte das implementações atuais de inteligência artificial, redes neurais profundas (*deep learning*), consiste em extrair padrões de grandes conjuntos de dados, origem de seu sucesso, mas igualmente de sua fragilidade. Durante anos diversas bases de dados tendenciosas foram usadas para desenvolver e treinar algoritmos de ia, sem nenhum escrutínio. “Reunir dados de qualidade em grande escala é caro e difícil. Criar grandes conjuntos de dados logo se tornou a versão da ia para a fabricação de salsichas: tedioso e difícil, com alto risco de usar ingredientes ruins”, pondera Hutson.

O ImageNet, banco de imagem utilizado intensamente pelos desenvolvedores de inteligência artificial, demorou uma década para reconhecer o viés na rotulagem de suas imagens, e, mesmo assim, só o fez após denúncia do artista norte-americano Trevor Paglen. Por meio do aplicativo ImageNet Roulette, idealizado por Paglen, qualquer usuário pode checar como o sistema do ImageNet classifica sua foto, basta fazer o *upload* da imagem no aplicativo (com muitas surpresas).

Os administradores do ImageNet, criado em 2009, publicaram um artigo, em janeiro de 2020, no qual reconheciam que, na categoria “pessoas”, 56% “são rótulos potencialmente ofensivos que não devem ser usados no contexto de um conjunto de dados de reconhecimento de imagem”. No final do processo de revisão, permaneceram apenas 6% das categorias originais de pessoas.²⁸

Brian Christian, no livro *The Alignment Problem: Machine Learning and Human Values* (*Problema do alinhamento: aprendizado de máquina e valores humanos*), cita o caso do banco de dados Labeled Faces in the Wild (lfw), criado em 2007 com base em artigos de notícias *online* e rotulado por uma equipe da UMass Amherst (Universidade de Massachusetts).²⁹ Em 2014, Hu Han e Anil

Jain, pesquisadores da Universidade Estadual do Michigan, notaram que nesse banco de dados mais de 77% das imagens eram de homens e, desse conjunto, mais de 83% de homens de pele clara. O ex-presidente dos Estados Unidos George W. Bush, dada sua visibilidade à época, tinha 530 imagens exclusivas, mais do que o dobro do conjunto de imagens de todas as mulheres de pele escura combinadas. Cinco anos depois, e 12 da data de constituição do Ifw, seus gestores postaram um aviso de isenção de responsabilidade, alertando que muitos grupos não estavam bem representados.

A agência governamental norte-americana Office of the Director of National Intelligence – supervisora da implementação do Programa de Inteligência Nacional, principal assessora do presidente, do Conselho de Segurança Nacional e do Conselho de Segurança Interna para assuntos de inteligência relacionados à segurança nacional – lançou em 2015 um banco de dados de imagens faciais denominado ijb-a, supostamente contemplando a diversidade da população norte-americana (etnia, gênero e raça). Entretanto, estudo das pesquisadoras Timnit Gebru e Joy Buolamwini, protagonistas do documentário da Netflix *Coded Bias*, constatou que 75% eram imagens de homens e, desse conjunto de dados, 80% de homens de pele clara, contra apenas 4,4% de imagens de mulheres de pele escura.

O problema do viés nos resultados dos sistemas de inteligência artificial, gradativamente, tem sensibilizado a sociedade. São múltiplas as origens do viés, desde a geração dos dados até as escolhas dos desenvolvedores, com forte contribuição de bases de dados tendenciosas. Nos últimos dois anos, por conta das denúncias, várias bases de dados, antes disponibilizadas na internet, foram suprimidas. Em junho de 2019, por exemplo, pesquisadores da Universidade Duke retiraram seu conjunto de dados Dukemtmc, formado por imagens coletadas de vídeos, principalmente de estudantes, gravados em um cruzamento movimentado do campus durante 14 horas em um dia de 2014. No período de três anos, esse banco de dados imperfeito foi utilizado por dezenas de empresas e agências governamentais em projetos de reconhecimento facial. Estudo recente da Universidade de Princeton constatou, inclusive, que versões do Dukemtmc foram usadas e citadas centenas de vezes em pesquisas, mesmo após ter sido “retirado do ar”.

No mesmo período, ficou inacessível o Diversity in Faces, base de dados de mais de um milhão de imagens faciais coletadas da internet, disponibilizada no início de 2019 por uma equipe de pesquisadores da ibm. “Ao todo, cerca de uma dúzia de conjuntos de dados de ia desapareceram – limpos às pressas por seus criadores depois que pesquisadores, ativistas e jornalistas expuseram uma série de problemas com os dados e as formas como eram usados, desde privacidade, preconceito racial e de gênero, até questões de direitos humanos”, segundo John Mcquaid, jornalista dedicado aos problemas da disseminação rápida e desregulada da inteligência artificial e vencedor do Prêmio Pulitzer.³⁰

Com base em investigação junto a cientistas de inteligência artificial, o jornalista Charles Q. Choi identificou algumas potenciais falhas desses sistemas, entre outras, a fragilidade na interpretação de fotos: alterar um único *pixel* em uma imagem pode fazer um sistema de reconhecimento de imagem “pensar que um cavalo é um sapo”.³¹ Outras falhas apontadas por Choi: viés, opacidade/“caixa-preta”, variável de incerteza intrínseca a modelos estatísticos de probabilidade e ausência de senso comum, ou seja, a incapacidade dos modelos de inteligência artificial de emitir conclusões lógicas com base em experiências do cotidiano (supostamente, como os seres humanos).

A aposta é que os modelos de ia tomem decisões de forma mais imparcial e mais transparente do que os seres humanos, mas, para essa aposta se concretizar, ainda há um longo caminho a percorrer. Enquanto os cientistas de inteligência artificial não descobrem como detectar e eliminar o viés dos resultados desses modelos, cabe à sociedade evitar a discriminação automatizada em massa e os imensos riscos aos seus usuários. Os sistemas de ia, particularmente em setores sensíveis, como educação e saúde, devem ser previamente auditados por agências reguladoras, com especial atenção às bases de dados.



IA e mercado de trabalho

A crescente adoção de tecnologias digitais (“transformação digital”) pelas instituições, públicas e privadas, está transformando tarefas, empregos e habilidades. Uma parcela não desprezível dos empregos recém-criados será, na próxima década, em ocupações totalmente novas ou ocupações existentes com inéditos conteúdos e requisitos de competências. Esse conjunto resultante de profissões emergentes reflete a adoção de novas tecnologias e a crescente demanda por novos produtos e serviços, impulsionadores de inéditos empregos na Economia Verde, na vanguarda da Economia de Dados e ia, na Economia do Cuidado, bem como novas funções na engenharia, na computação em nuvem, em marketing/vendas e na produção de conteúdo.

A experiência da pandemia da covid-19 forçou as instituições a anteciparem o “futuro do trabalho”, implementando tecnologias de automação anteriormente associadas a projetos de longo prazo. A aceleração da digitalização, consequentemente da automação, impacta o mercado de trabalho em duas frentes: carência de mão de obra qualificada e deslocamento de trabalhadores.

Entre as medidas para mitigar as consequências negativas da automação “inteligente”, baseada nas tecnologias de ia, são imperiosos os investimentos em educação para qualificar e requalificar os trabalhadores, estando eles ou não empregados. A educação pode evitar que parte significativa da sociedade seja desconectada dos benefícios gerados pelo avanço tecnológico.

A velocidade e o ritmo da transição serão influenciados pelas condições da economia local, das políticas públicas e da capacidade de cada setor e instituição em acompanhar as mudanças em curso na economia global. A tendência é aumentar a desigualdade entre os países, entre as empresas, particularmente entre as micro/pequenas empresas e as médias/grandes empresas, e entre os indivíduos. O grau de desigualdade é proporcional ao acesso a recursos pertinentes.

Este bloco tem seis artigos, e cada um deles aborda aspectos específicos dos desafios atuais para evitar a classe dos “inempregáveis” ou “inúteis”, como designa Yuval Harari os cidadãos que não apenas serão desempregados, mas também não serão empregáveis no futuro do trabalho.

Você está preparado para trabalhar no século xxi?

12.7.2019

Estudos de consultorias e instituições internacionais sobre o mercado de trabalho divergem quanto aos números, porque têm base em metodologias distintas, contudo, convergem sobre a tendência: eliminação crescente de funções, ameaça aos empregos.

No Brasil temos dois estudos: da Universidade de Brasília, que indica que 54% das funções no Brasil têm probabilidade de ser eliminadas até 2026; e do Instituto de Pesquisa Econômica Aplicada (Ipea), órgão ligado ao Ministério do Planejamento, que indica que mais de 50% das funções serão eliminadas até 2050, ou seja, 35 milhões de trabalhadores formais correm risco de perder seus empregos para a automação. Por outro lado, com 13% de taxa de desemprego, as empresas enfrentam dificuldade de preencher vagas em aberto por falta de candidatos qualificados.

As projeções são de um mercado de trabalho cada vez mais desigual, com as funções de alto desempenho extremamente lucrativas, e as demais com perdas salariais, ou eliminadas pela automação. O grupo de elite norte-americano, por exemplo, quase dobrou sua participação na renda nacional entre 1980-2016; em 2017, o 1% mais rico dos norte-americanos equivalia a quase o dobro de riqueza dos 90% mais pobres.

A Uber ilustra bem o que está por vir. O número de motoristas cadastrados cresceu em 50% entre 2016-2018 (de 50 milhões para 100 milhões). No Brasil são cerca de 600 mil motoristas; o pleno sucesso de seu projeto de carro autônomo, em teste em várias cidades, gerará um lucro extraordinário aos seus acionistas e uma perda total para os seus motoristas. A automação inteligente vai invadir o varejo, as transportadoras, os bancos e uma infinidade de funções em quase todos os setores, atingindo fortemente a classe média.

Ao contrário de processos associados a tecnologias disruptivas anteriores, os novos modelos de negócio não são intensivos em mão de obra. Na automação das fábricas no século xx, por exemplo, os trabalhadores dispensados tinham como alternativa o setor de serviços, em plena expansão. A Economia de Dados não oferece muitas alternativas. A montadora gm demorou 70 anos para gerar um lucro de 11 bilhões de dólares com 840 mil funcionários, e o Google precisou de meros 14 anos para lucrar 14 bilhões de dólares com 38 mil funcionários. O exemplo talvez mais emblemático: em 2012 a Kodak abriu falência com 19 mil funcionários, tendo chegado a empregar 145 mil; no mesmo ano, o Instagram foi comprado pelo Facebook por 1 bilhão de dólares com apenas 13 funcionários.

Na competição entre o trabalhador humano e o “trabalhador-máquina”, os humanos estão em desvantagem: a) a manutenção é mais barata, as máquinas trabalham quase que em moto-contínuo (sem descanso, sem férias, sem doenças), com um custo médio menor por hora trabalhada (49 dólares na Alemanha e 36 nos Estados Unidos, contra 4 dólares do “robô”); e b) as máquinas inteligentes se aperfeiçoam automática e continuamente, e o custo de reproduzi-las é significativamente menor do que o custo de treinar profissionais humanos para as mesmas funções.

As transformações no mercado de trabalho não advêm exclusivamente da automação inteligente, mas igualmente de novas configurações como *home office* e contratação por projeto (“pejotização”). Outro fator é a categoria chamada *gig economy* – plataformas e aplicativos *online*, *freelancers*; os aplicativos Uber, iFood, Rappi e 99 são hoje o maior empregador do país, com cerca de 3,8 milhões de trabalhadores, representando 17% dos 23,8 milhões de trabalhadores autônomos, segundo o Instituto Brasileiro de Geografia e Estatística (ibge). A tendência é as empresas reduzirem o número de empregados fixos, regidos pelas leis trabalhistas como a clt, com redução de custos e ganhos de eficiência, inclusive na qualidade do serviço prestado.

As profissões-chave no mercado de trabalho dos próximos anos, segundo as consultorias especializadas, são analista de dados, cientista de dados, desenvolvedor de *software* e aplicativos, especialista em comércio eletrônico, especialista em mídias sociais, profissional de ia com ênfase em aprendizado de máquina, especialista em *big data*, analista de segurança da informação e engenheiro de robótica. Em paralelo, existe um grande potencial em funções centradas em habilidades humanas, como atendimento ao cliente, vendas e marketing, treinamento e desenvolvimento de pessoas e cultura, gestão da inovação e desenvolvimento organizacional. Até 2030, mais de um terço das habilidades essenciais para a maioria das ocupações será composto por habilidades que ainda não são cruciais para o trabalho atual.

Para não perder a relevância econômica e social no século xxi, o desafio é identificar quais as habilidades necessárias para que um robô não roube seu emprego, e se capacitar. Lição de casa: liste todas as funções/tarefas desempenhadas no seu trabalho, agrupe em colunas as mais suscetíveis à automação e as que requerem habilidades ainda exclusivamente humanas e prepare-se para desempenhar melhor estas últimas.

Inteligência artificial pode democratizar o acesso à Justiça?

13.12.2019

Em 2015, aos 19 anos, o programador Joshua Browder criou um sistema de inteligência artificial do tipo *chatbot* que, a partir do diálogo com o usuário, orienta sobre a melhor forma de contestar multas de estacionamento, gerando automaticamente os documentos legais apropriados. Com base na tecnologia Watson-ibm, o DoNotPay – disponível em Nova York e no Reino Unido – é um serviço jurídico *pro bono* (totalmente gratuito). De 2015 a 2017, foram atendidos 250 mil casos, anulando cerca de 4 milhões de dólares em multas. Trata-se de mera ilustração da transformação em curso na advocacia.

Para dar conta do volume de documentos envolvidos nos processos jurídicos, em 2012, surgiu a *technology-assisted review* (tar), considerada a primeira aplicação de ia na prática legal. Sua função é organizar, analisar e pesquisar conjuntos de dados com eficiência 50 vezes maior que a dos métodos tradicionais, em rapidez e precisão, reduzindo custos e tempo de revisão da *electronically stored information* (esi).

O primeiro “robô-advogado” foi o Ross. Convergindo tecnologias de pesquisa *online* com processamento de linguagem natural da Watson-ibm, Ross lê mais de um milhão de páginas legais por minuto. Projetado por uma equipe diversificada de profissionais, seu propósito é “mudar o mundo, oferecendo assistência prática à grande maioria das pessoas em risco legal que não têm acesso à justiça”.³² No momento, está disponível somente nos Estados Unidos (em breve, globalmente).

Como forma de testar os sistemas inteligentes, em 2018, 20 experientes advogados norte-americanos (com passagem pelo banco Goldman Sachs, pela Cisco e por grandes escritórios de advocacia) foram confrontados com o algoritmo de inteligência artificial LawGeex. O desafio era rever cinco *non-disclosure agreements* (ndas, acordos de confidencialidade), identificando os riscos em cada contrato legal. O sistema de ia alcançou 94% de precisão *versus* 85% dos advogados (na média: 94% para advogados com maior desempenho, e 67% para advogados com menor desempenho), com uma diferença de tempo de 26 segundos para a ia contra 92 minutos para os advogados!

Esses avanços, como nos demais setores da economia, trazem enormes benefícios à sociedade, pela redução de custos e pelo aumento de eficiência (maior precisão, menor tempo); ademais, democratizam o acesso à defesa legal de um contingente majoritário da população que não tem condições de arcar com os custos da advocacia tradicional.

Simultaneamente, contudo, esses mesmos avanços têm efeitos negativos sobre o mercado de trabalho; segundo a consultoria McKinsey, 23% do trabalho atual do advogado médio já pode ser substituído por sistemas inteligentes. Esse cenário afeta particularmente os jovens advogados: a automação inteligente tende a eliminar as tarefas rotineiras na prática legal, funções em geral exercidas por jovens associados, ou seja, os candidatos ao primeiro emprego. No Brasil o cenário é mais preocupante: são cerca de 1,2 milhão de advogados no país, 1.210 cursos de direito e 900 mil estudantes (no resto do mundo são 1.100 cursos de direito, menos do que o total do país).

Indicadores das economias desenvolvidas dão uma noção do que está por vir: segundo a Deloitte, companhia de auditoria e consultoria empresarial, 100 mil empregos na área advocatícia serão eliminados no Reino Unido até 2025; o banco norte-americano jpm automatizou 360 mil horas de trabalho de advogado no último ano; mais de 36% dos escritórios de advocacia norte-americanos, e mais de 90% dos grandes escritórios de advocacia do mundo (escritórios com mais de mil advogados), estão usando ativamente a inteligência artificial. No Brasil, estamos nos primórdios: de acordo com Alexandre Zavaglia, presidente da New Law School, que oferece cursos sobre as transformações tecnológicas no universo jurídico, apenas de 2 a 4% do meio jurídico está usando ia.

No campo da ética, emergem inúmeras questões a serem equacionadas: como garantir a precisão, a legalidade e a justiça das decisões de inteligência artificial? Os advogados serão acusados de negligência por confiarem em sistemas inteligentes que cometem erros? Ou, ao contrário, serão acusados de negligência se não fizerem uso de ia que exceda as capacidades humanas em certas tarefas?

O futuro da advocacia foi um dos temas do Primeiro Seminário Novas Tecnologias e Sistemas de Inteligência Artificial, promovido pelo Tribunal Regional Federal da 3ª Região, em 6 de dezembro. Os palestrantes – desembargadores, juízes, conselheiros, auditores e acadêmicos – se dividiram entre alertar sobre a iminente ameaça de desaparecimento da profissão do advogado até a extinção da própria Justiça como a conhecemos, e o reconhecimento da inexorabilidade da adoção das tecnologias de inteligência artificial (“porque é o que as pessoas querem”). A unanimidade foi sobre a premência de atualizar os profissionais do setor, advogados e juízes, e adequar as grades curriculares dos cursos de Direito ao novo mercado. Um ponto de partida é acompanhar o site Artificial Lawyer (www.artificiallawyer.com).

O Brasil tem atualmente 250 mil vagas em aberto, ou seja, ofertas de emprego sem candidatos qualificados, e esse número tende a aumentar nos próximos anos, segundo a Associação das Empresas de Tecnologia da Informação e Comunicação (Brasscom). Cinco em cada 10 empresas industriais brasileiras enfrentam dificuldades com a falta de trabalhadores qualificados, de acordo com a Confederação Nacional da Indústria (cni).

Por outro lado, segundo o ibge, a taxa de desemprego bateu 11% em dezembro de 2019 (11,6 milhões de brasileiros, cerca de 15% da população economicamente ativa), índice subestimado, porque não leva em conta os que desistiram de procurar emprego, os que estão empreendendo para subsistir – 50% do total dos empreendedores faturam aproximadamente mil reais por mês, e apenas 1% dos empreendedores iniciais e 3,2% dos estabelecidos faturam acima de 5 mil reais por mês (dados do Global Entrepreneurship Monitor, gem, Brasil/2017) –, sem falar nos 41,1% do total da população ocupada que trabalham na informalidade (sem carteira assinada). É fato que as novas tecnologias estão gerando, simultaneamente, desemprego e oferta de vagas em aberto, além de desequilíbrio salarial: salários crescentes para funções qualificadas e decrescentes para as demais funções.

A combinação entre os avanços da inteligência artificial e da robótica, se por um lado acelera a produtividade e o crescimento, com a redução de custos e o aumento da eficiência, por outro tem o potencial de gerar desemprego e desigualdade (produz efeito negativo sobre a renda ao aumentar a competição pelos empregos remanescentes). Como alerta a Organização Internacional do Trabalho (oit), no curto/médio prazo as funções mais ameaçadas são as de menor qualificação, em geral, exercidas pela população de baixa e média renda (proporcionalmente em maior número).

A consultoria McKinsey Digital, no artigo *The Impact of Automation on Employment for Women and Minorities* (2019), atesta a vulnerabilidade, por exemplo, dos trabalhadores afro-americanos concentrados nas atividades mais propensas à automação: super-representados na categoria motoristas de caminhão, com 80% das horas de trabalho ameaçadas pelos caminhões autônomos, e sub-representados na categoria desenvolvedores de *software*, com 15% de probabilidade de automação.

O aplicativo Chronicle, do *New York Times*, mostra que, entre 2009 e 2016, multiplicou-se por 10 a frequência da palavra “desigualdade” em suas matérias,

atingindo a proporção de uma a cada 73 palavras (não há dados mais recentes); supondo uma tendência geral, pode ser um indicador da crescente visibilidade do tema na mídia. Os Estados Unidos, país das estatísticas, têm atualmente a menor taxa de desemprego em 50 anos, 3,5% da população, mas, ao mesmo tempo, aumentou a desigualdade: a elite norte-americana quase dobrou sua participação na renda nacional entre 1980 e 2016; em 2017, o 1% mais rico dos norte-americanos equivalia a quase o dobro de riqueza dos 90% mais pobres.

O desequilíbrio do mercado de trabalho deve-se, em parte, ao fato de que as novas funções não estão substituindo proporcionalmente as funções eliminadas, basicamente por duas razões: primeiramente porque é maior o número de funções repetitivas e previsíveis, foco da substituição por máquinas inteligentes, em comparação com as novas funções ofertadas; o estímulo à transformação digital é reduzir custos e aumentar a eficiência, processos intensivos em tecnologia (e não em mão de obra). Em segundo lugar porque é um equívoco considerar as habilidades listadas pelos relatórios de consultorias, e enfatizadas por muitos “especialistas” – raciocínio lógico, empatia, pensamento crítico, compreensão de leitura, argumentação, comunicação clara e persuasiva, discernimento, bom senso, capacidade de tomar decisão, aprendizagem ativa, fluência de ideias, originalidade –, como inerentes aos seres humanos. Elas são potencialmente habilidades humanas, representam potencial vantagem comparativa dos trabalhadores humanos. Para florescerem, contudo, dependem de condições apropriadas, e essas condições apropriadas não estão disponíveis para a maioria da população nos países desenvolvidos e, particularmente, nos países em desenvolvimento. Responda sinceramente: você possui todas essas “habilidades humanas”? Quantas pessoas você conhece que as possuem?

Os relatórios internacionais, frequentemente citados como fontes, devem ser lidos com cautela. O relatório de janeiro de 2020, por exemplo, do Fórum Econômico Mundial (WEF) lista várias profissões como novas, quando apenas receberam uma nova nomenclatura, e outras requerem as “famosas habilidades humanas”. O relatório projeta que até 2022, 37% das oportunidades estarão na Economia do Cuidado, ou seja, em parte refere-se aos “cuidadores”: a função que mais cresce atualmente nos Estados Unidos (longevidade da população, mudança na estrutura familiar, entre outros fatores), mas com curva declinante de remuneração.

As demais funções com altas taxas de crescimento incluem especialistas em inteligência artificial, cientistas de dados, engenheiros de computação, analista de dados; com baixas taxas de crescimento incluem técnicos e assistentes em geral. Os autores do relatório do WEF, inclusive, relativizam as próprias projeções: “O crescimento e a escala absoluta de várias profissões serão determinadas de maneira distinta pelas atuais escolhas e investimentos feitos pelos governos”.

No Brasil, a precariedade do ensino fundamental, perpetuada na baixa qualidade da maioria dos cursos superiores, é uma barreira à formação de

profissionais adequados ao mercado de trabalho do século xxi. Abstraindo as exceções, a mobilidade no emprego é prerrogativa da elite, com acesso a formação, e não apenas a treinamento. Ou seja, quem “se salva” são os profissionais de alta qualificação, que representam menos de 1% da população brasileira. Segundo estudo de pesquisadores do Fundo Monetário Internacional (fmi), o grau de preocupação com a ameaça tecnológica aos empregos, não por acaso, está correlacionada aos níveis de educação e ao acesso à informação.³³

Até que provem o contrário com dados consistentes, o impacto no mercado de trabalho é um dos efeitos sociais mais perversos da adoção da inteligência artificial. As empresas de tecnologia negam as evidências temerosas de serem responsabilizadas pelas externalidades negativas, e os relatórios de consultorias e organismos internacionais, se lidos atentamente, não são conclusivos.

A construção de um futuro promissor depende de política pública, que por sua vez depende de um Estado competente e saudável, com capacidade de promover ações efetivas e coordenadas em larga escala, de formar ecossistemas com o setor público, as universidades e centros de pesquisa, e com os investidores (e não apenas criar arcabouço regulatório). Sem política pública, dificilmente o Brasil entrará efetivamente no século xxi.

Como evitar a classe dos “inempregáveis”?

23.10.2020

A série de televisão *Os Jetsons*, criada pelo estúdio Hanna-Barbera em 1962 e exibida no Brasil em 1963 pela extinta tv Excelsior, ao retratar o cotidiano de uma família de classe média no ano 2062, introduziu no imaginário popular a ideia ficcional do futuro: carros voadores, automação, robôs domésticos, cidades suspensas. O historiador Michael Bess, no livro *Our Grandchildren Redesigned* (*Nossos netos redesenhados*), pondera que a série reproduz um equívoco recorrente: a tendência de imaginar que as tecnologias evoluirão radicalmente, enquanto nós, humanos, permaneceremos fundamentalmente os mesmos.³⁴

Sucessos como as franquias *Star Wars* e *Star Trek*, e filmes de ficção científica como *Blade Runner*, de Ridley Scott, e *AI*, de Steven Spielberg, retratam espécies alienígenas e robôs inteligentes contracenando com humanos iguais; quando modificados, como os personagens de filmes como *A mulher biônica* e *Homem-Aranha*, são por acidentes. No campo da ficção científica, a única aparente exceção são os seres humanos geneticamente aprimorados da obra de Aldous Huxley *Admirável mundo novo* (1932).

Difícil prever como seremos no futuro, mas é fato que os corpos e mentes humanos já estão sofrendo intervenções causadas por drogas medicinais e nanotecnologias. Contudo, enquanto aguardamos a reengenharia radical do *Homo sapiens* – reescrita dos códigos genéticos, alteração do equilíbrio bioquímico, fusão com dispositivos não orgânicos, conexão direta entre cérebros e máquinas –, o desafio é evitar cair na classe dos “inempregáveis”, como alerta Yuval Harari.³⁵

O cientista da computação e investidor Kai-Fu Lee observa que as revoluções industriais anteriores desqualificaram o trabalho (*deskilling*).³⁶ As linhas de montagem das fábricas transformaram as tarefas que eram feitas por pessoas qualificadas, sapateiro-artesão por exemplo, em linhas de montagem com trabalhadores com baixa qualificação (*nonskilled labor*), em que cada um adiciona uma diminuta parte ao todo, cenário retratado com rigor por Joshua B. Freeman no livro *Mastodontes*.³⁷ A revolução 4.0, ao contrário, requer o chamado *upskilling*, ou *reskilling*, ou seja, demanda a qualificação do trabalho para desempenhar tarefas mais complexas e multidimensionais (as tarefas rotineiras, repetitivas, previsíveis aos poucos estão sendo desempenhadas pela automação inteligente, ou seja, tecnologias de inteligência artificial).

Paradoxalmente, num mundo dominado pela tecnologia, as ciências humanas têm papel de destaque, na medida em que o profissional do futuro vai

lidar com questões que exigem, além de conhecimentos técnicos, habilidades de lógica, análise crítica, empatia, comunicação e design. Scott Hartley defende a parceria entre as ciências exatas e as humanas, em que as primeiras focam no “como fazer” da revolução tecnológica, e as segundas, no “por quê”, “para quê” e “quando”.³⁸ Vale atentar, e tomar como referência, que os empreendedores do Vale do Silício, em geral, têm formação em ciências humanas e sociais e reconhecem que essas habilidades foram determinantes para o sucesso de seus projetos. Além disso, parte deles matriculam seus filhos em escolas “humanistas”.

A pergunta de um milhão de dólares é quais profissões serão eliminadas e quais serão as profissões do futuro. Previsões sobre o mercado de trabalho divergem por conta de metodologias distintas, mas convergem sobre a tendência: em sua maioria, potencial desemprego massivo. Estudo do Fórum Econômico Mundial (wef), *The Future of Jobs Report 2020*, em parceria com LinkedIn, Coursera, FutureFit AI e adp, constata que a automação está aumentando mais rapidamente do que o previsto.³⁹ Com a covid-19, cerca de 84% dos executivos consultados aceleraram seus planos de digitalização e adoção de novas tecnologias. As lacunas de competências continuam a ser altas; em média, as empresas estimam que cerca de 50% dos trabalhadores precisarão de requalificação. “Na ausência de esforços proativos, a desigualdade provavelmente será exacerbada pelo duplo impacto da tecnologia e da recessão pandêmica”, vaticina o estudo.

O wef, sensível à necessidade de preparar o futuro, lançou a plataforma The Reskilling Revolution, que convida líderes e organizações a contribuir em torno de cinco eixos. O propósito é qualificar e requalificar a força de trabalho, fornecendo às empresas e economias mão de obra adequada para exercer as novas funções. Vários países e empresas já aderiram, parcerias foram firmadas, entre elas com a empresa de tecnologia educacional norte-americana Coursera, o LinkedIn, a Salesforce e a pwc.

A atuação do wef é respaldada localmente pelas políticas públicas dos países, particularmente os países líderes em transformação digital que tratam a educação como estratégia para o desenvolvimento econômico. Nos Estados Unidos, por exemplo, o governo convocou o setor privado a se comprometer com a qualificação e requalificação de sua força de trabalho por meio do Pledge to America’s Workers: mais de 415 empresas do setor privado já se comprometeram com 14,5 milhões de oportunidades de aprimoramento de carreira nos próximos cinco anos. Na França, o aplicativo móvel Mon Compte Formation é dedicado a integrar a formação profissional e a aprendizagem ao longo da vida. “Esses esforços combinados do setor privado e dos governos podem catalisar melhor educação, habilidades e empregos para apoiar um bilhão de pessoas e servir como exemplos globais”, pondera o wef.

No Brasil não temos política pública nem ecossistema favorável, a perspectiva é individual, ou seja, cada um assume o protagonismo de sua própria

carreira. A boa notícia é que informações e conhecimento estão disponíveis. Há oferta de cursos variados, presenciais e *online*, alguns a custos razoáveis e outros gratuitos. Proliferam publicações na mídia especializada e nos grandes veículos de comunicação. Com a covid-19, as *lives* invadiram nosso cotidiano. Gradativamente, junto à familiaridade com os temas, virá a capacidade de discernir os bons conteúdos (curadoria). O estudo do *wef* indicou um aumento de quatro vezes no número de indivíduos que procuram aprendizagem *online* por iniciativa própria.

Há dois conceitos-chave a serem observados: um é o *lifelong learning*, aprendizagem ao longo da vida, fundamental para acompanhar a aceleração atual, requerendo atualização contínua que extrapola o ensino formal; outro é a *autorregulação da aprendizagem*, em que cada um estrutura, monitora e avalia seu próprio aprendizado, ampliando sua capacidade de retenção de conteúdo e engajamento. O desafio é montar um programa de aprendizado eficiente. O futuro será de quem “aprende a aprender”.

Para não perder a relevância econômica e social no século xxi, a recomendação é cada um identificar quais as habilidades necessárias para que o “robô” não roube seu emprego, e se capacitar. Lição de casa (sugerida em coluna anterior): liste todas as funções e tarefas desempenhadas no seu trabalho, agrupe em colunas as mais suscetíveis à automação e as que requerem habilidades ainda exclusivamente humanas, e prepare-se para desempenhar melhor estas últimas. Não é uma boa apostar na renda mínima (*ubi, universal basic income*); se e quando ela for concretizada, garantirá apenas a sobrevivência.

Inteligência artificial no jornalismo: ameaça ou oportunidade?

6.8.2021

Em maio de 2020, o jornal *The Guardian* anunciou que o site do msn, da Microsoft, acessado diariamente por milhões de britânicos, havia substituído parte de seus jornalistas por sistemas de inteligência artificial. Longe de ser um fato isolado, trata-se de uma tendência do ecossistema de mídia, como indica o relatório *Journalism, Media, and Technology Trends and Predictions 2021* (*Tendências e previsões para o jornalismo, a mídia e a tecnologia em 2021*), publicação do Instituto Reuters e da Universidade de Oxford.⁴⁰

Segundo o relatório, 69% dos entrevistados destacaram o protagonismo da inteligência artificial nos próximos anos – 18% apontaram a conectividade 5G – em curadoria, produção e distribuição de conteúdo. Para atender a diversidade de formatos de consumo de notícias, o relatório sinaliza a necessidade de modelos híbridos, que combinem mídia e tecnologias digitais, denominados de “*media tech*” (*media technology*). O termo se aplica a qualquer dispositivo ou tecnologia para criar, produzir, distribuir e gerenciar mídia, tais como rede de distribuição de conteúdo, mídia interativa, convergência digital, realidade aumentada, realidade mista, infraestrutura de mídia, entre outros.

O termo “*media tech*”, inclusive, foi utilizado pelo presidente executivo do Grupo Globo, Jorge Nóbrega, para definir o novo modelo de negócio do grupo após reestruturação. A disputa por publicidade com as *big techs*, em parte, tem afetado os resultados financeiros dos grupos de mídia tradicionais: o lucro da Globo, por exemplo, passou de 2,7 bilhões de reais, em 2010, para 0,1 bilhão, em 2020. Em contrapartida, a receita publicitária da Amazon cresceu 83% no segundo trimestre de 2021, comparativamente ao mesmo período do ano anterior, alavancada pela publicidade *online* de marcas e pequenos negócios no *marketplace*, nas plataformas de vídeo (Fire tv, imdb tv e Twitch), na plataforma de *streaming* de games e nos vídeos ao vivo.⁴¹ Em 2020, do investimento total em publicidade digital nos Estados Unidos, 63,5% foi direcionado para as três grandes plataformas: 29,8% para o Google; 23,5% para o Facebook; 10,2% para a Amazon. O faturamento em anúncios *online* do Google foi de 146,8 bilhões de dólares, e do Facebook, de 84,1 bilhões de dólares.

As organizações de mídia, gradativamente, estão se apropriando da inteligência artificial para mudar a maneira como as notícias são geradas, produzidas, publicadas e compartilhadas. Processo ainda embrionário, num futuro próximo a expectativa é que os sistemas inteligentes serão responsáveis por gerar parte dos textos. No momento, a função principal dos algoritmos de ia é

“varrer” e classificar publicações em vários canais (mídia, redes sociais, relatórios privados e públicos, comunicados, entre outros), ou seja, acelerar a pesquisa ao correlacionar, rápida e eficientemente, grandes conjuntos de dados por marcas semânticas e categorias (eventos, pessoas, locais, datas). A seção “Comentários” do *The New York Times*, por exemplo, é moderada por 14 jornalistas que revisam manualmente mais de 11 mil comentários diários, tarefa em vias de ser desempenhada pelo sistema de inteligência artificial criado pela Jigsaw/Alphabet, controladora do Google – reduz custos e aumenta a eficiência, motivações iniciais para a adoção da ia.⁴²

Em paralelo, cresce o movimento de entrega de conteúdo personalizado com base no perfil do leitor, movimento favorável aos anunciantes pelo potencial aumento da taxa de conversão, e conteúdo personalizado por perfil do jornalista: um especialista em finanças, por exemplo, recebe links para matérias e imagens relacionadas a esse tema (a inteligência artificial no papel de “assistente” do jornalista). Nesse caso, são múltiplas as experiências: a *Forbes* lançou o Bertie, sistema agregador de notícias e sugestão de conteúdo; o *Washington Post* lançou o Heliograf, capaz de gerar textos a partir de dados quantitativos, sistema já utilizado em 2016 na cobertura dos Jogos Olímpicos e das eleições norte-americanas; a Bloomberg está usando o Cyborg para criar e gerenciar conteúdo. As agências de notícias, como a Associated Press e a Reuters, seguem na mesma linha.

Em 2018, a Reuters lançou o Lynx Insight, sistema de inteligência artificial para analisar dados, sugerir temas para matérias e até mesmo escrever algumas frases. Segundo Reg Chua, editor-executivo de Operações Editoriais, Dados e Inovação da Reuters, o objetivo é maximizar o melhor das máquinas inteligentes (identificar padrões e fatos em grandes bases de dados) e das equipes humanas (fazer perguntas, avaliar relevância, entender o contexto). Provavelmente para minimizar o impacto negativo no mercado de trabalho e na oferta de empregos, Chua cometeu o equívoco de argumentar que o Lynx Insight seria tão revolucionário para um jornalista quanto o telefone, duas tecnologias de natureza absolutamente distintas.⁴³

O processo de escrita de um texto é chamado de “geração de linguagem natural” (*natural language generation*, nlg), a partir da definição inicial do formato desejado; o nlg já é usado, por exemplo, em relatórios de negócio, atualizações de portfólio financeiro e e-mails personalizados. Os modelos de nlg disponíveis são o Quill, da Narrative Science, Amazon Polly, WordSmith, da Automated Insights, e Google Text-to-Speech; algumas organizações criaram internamente seus modelos, como o já citado Heliograf, do *Washington Post*. A vantagem, além de lidar com grandes volumes de dados, é a capacidade de produzir textos em frações do tempo de um jornalista humano. O jornal *The New York Times*, recentemente, lançou um desafio para o leitor diferenciar um conteúdo escrito por um humano de um conteúdo escrito por um sistema de inteligência artificial.⁴⁴

A conferência *Artificial Intelligence and the Future of Journalism: Will Artificial Intelligence Take Hold of the Fourth Estate?* (*Inteligência artificial e o futuro do jornalismo: a inteligência artificial vai dominar o quarto poder?*), organizada, em maio 2021, pela Federação Europeia de Jornalistas (efj), debateu o dilema de se a ia é uma ameaça ou uma oportunidade para o setor de mídia. Na perspectiva do presidente da efj, Mogens Blicher Bjerregaard, as questões mais urgentes a serem enfrentadas são: a adoção da ia potencializa o risco de ampliar a lacuna entre a grande e a pequena mídia; a alfabetização dos jornalistas em dados, enfatizando a urgência de qualificar e requalificar para as novas habilidades; e os desafios éticos.

As empresas de mídia estão apostando na inteligência artificial como meio de oferecer experiências personalizadas e melhorar a eficiência da produção. Como em qualquer setor, é importante estabelecer diretrizes e arcabouços regulatórios, ainda mais pelo papel social do jornalismo: notícia de qualidade é do interesse de toda a sociedade.

A sustentabilidade do trabalho depende da coexistência positiva de humanos e IA

18.3.2022

A Amazon decidiu, em 2022, dobrar o teto salarial (anual) dos funcionários de tecnologia de 160 mil para 350 mil dólares. Segundo o comunicado interno, a mudança visa alinhar a Amazon com as gigantes de tecnologia como Google, Facebook, Apple e Microsoft, enfrentando a intensa competitividade do mercado de trabalho em 2021 para reter e recrutar talentos. O movimento da Amazon ilustra o protagonismo da tecnologia num cenário de automação acelerada.

É cada vez mais intenso o debate público sobre o futuro do trabalho, particularmente os efeitos das mudanças tecnológicas sobre o emprego, os salários e a desigualdade. Com o propósito explícito de “agregar cientificidade” ao debate, o Israel Public Policy Institute (IPPI) publicou o artigo “Race Against the Machine? The Role of Technological Change for Employment, Wages and Inequality”⁴⁵ de autoria de Ulrich Zierahn, professor da Universidade de Utrecht, Holanda. O ponto de partida de Zierahn é o conceito de “Routine Replacing Technological Change” (RRTC), introduzido pelo professor de economia do MIT David Autor.

Autor, um dos maiores especialistas em automação do trabalho, ao lançar o conceito de RRTC em 2003⁴⁶ já alertava que: a) as mudanças tecnológicas alteram as habilidades profissionais; e b) a tendência é o “capital computacional” substituir os trabalhadores na execução de tarefas manuais e cognitivas, inclusive as mais complexas. O avanço da inteligência artificial (IA), com a chamada “automação inteligente”, acentuou o processo de automação em curso desde meados do século XX. A principal vantagem da IA sobre os trabalhadores humanos é a sua capacidade de detectar padrões “invisíveis” em grandes conjuntos de dados (*big data*), gerando previsões mais assertivas, consequentemente, melhores decisões, além de permitir que os sistemas “aprendam” com os dados num processo de aperfeiçoamento contínuo.

Como reconhece o Fundo Monetário Internacional (FMI),⁴⁷ não há consenso em torno da premissa de que a automação gera crescimento e desigualdade por parte de economistas e estudiosos das novas tecnologias. O FMI identifica duas perspectivas: a) os pessimistas da tecnologia que temem uma distopia econômica de extrema desigualdade e conflito de classes com previsões de queda acentuada da taxa de emprego; e b) os otimistas da tecnologia que, mesmo reconhecendo os impactos negativos da automação a curto prazo, baseiam-se nos processos históricos anteriores de mudança tecnológica com vetor positivo entre destruição e criação de empregos, com aumento de salários e de

renda per capita. Em qualquer cenário, mantidas as condições atuais, a automação é positiva para o crescimento econômico e negativa para a desigualdade. Kai-Fu Lee adverte que “o século XXI pode trazer um novo sistema de castas, dividido em uma elite plutocrática de IA e as massas em lutas impotentes”.⁴⁸

A automação inteligente incide mais fortemente sobre os empregos de salário médio, polarizando o trabalho entre empregos de baixa e alta renda. Em paralelo, a substituição do trabalhador humano pelos sistemas inteligentes gera efeito negativo sobre a renda ao aumentar a competição pelos empregos remanescentes (redução salarial). Segundo o Bureau of Labor Statistics dos EUA,⁴⁹ por exemplo, as duas profissões que mais crescem no país são os auxiliares de saúde domiciliar e os auxiliares de cuidados pessoais, com curva salarial decrescente.

As projeções sobre o futuro do trabalho estão permeadas de imprecisões metodológicas e/ou interpretativas, inclusive o famoso estudo dos pesquisadores britânicos Carl Benedikt Frey e Michael Osborne.⁵⁰ O elemento sensível em qualquer pesquisa é a metodologia, e, nesse caso, os autores extraíram suas projeções de expectativas de especialistas sobre quais funções poderiam ser automatizadas, desconsiderando fatores críticos que extrapolam a tecnologia, por exemplo, mudança cultural e processual nas organizações, capacidade de investimento, arcabouço regulatório, além da conjuntura política.

As transformações na economia afetam diretamente o ritmo, a intensidade e a configuração da automação, consequentemente, o presente e o futuro do trabalho. Dani Rodrik, professor da John F. Kennedy School of Government da Universidade Harvard, crê que a economia global após a crise de 2008, a pandemia da covid-19 e a guerra da Ucrânia será mais fragmentada e regionalizada, decretando o fim da hiperglobalização.⁵¹ Por outro lado, a sociedade hiperconectada gera um conjunto de dados extraordinários, estimulando a proliferação de modelos de negócio baseados em dados (*data-driven business models*). São múltiplos os exemplos que recomendam não hipervalorizar, e isolar, os efeitos da tecnologia. É inexorável, contudo, que a sustentabilidade do trabalho como o concebemos está ameaçada.

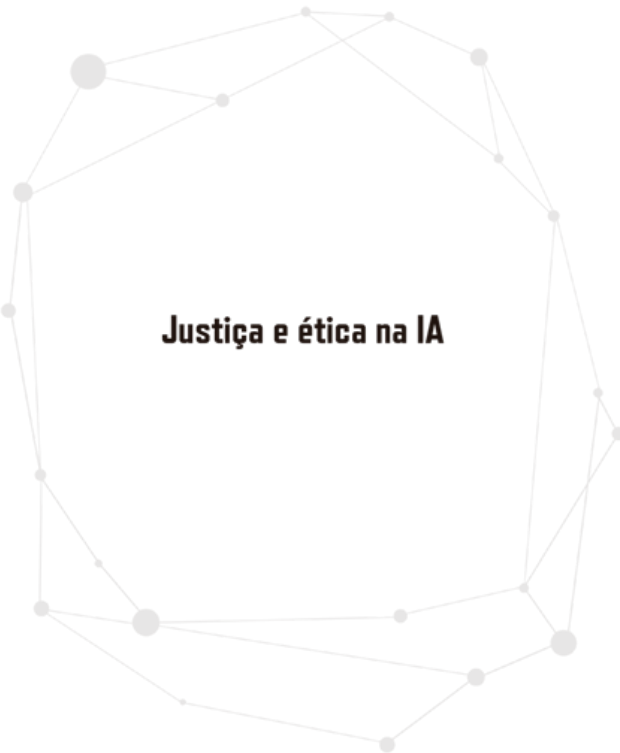
A pandemia, ao acelerar a digitalização, impactou o trabalho em duas frentes: a) a escassez de mão de obra qualificada – 68% dos executivos brasileiros alegam dificuldade para encontrar profissional qualificado para posições-chave, índice superior ao registrado em países da região, como Argentina (40%), Costa Rica (40%) e México (38%) –; e b) a dificuldade de os trabalhadores conseguirem uma ocupação. No Brasil, a pandemia ampliou a desigualdade entre a educação pública e a privada, sinalizando que os jovens menos favorecidos cheguem ao mercado de trabalho apresentando deficiências de formação que nem sempre as empresas estão dispostas a suprir em seus programas de treinamento e/ou qualificação. O relatório do Banco Mundial “Employment in Crisis: The Path

to Better Jobs in a Post-COVID-19 Latin America”,⁵² publicado em 20 de julho de 2021, prevê que a crise da covid-19 trará efeitos duradouros sobre o emprego, sendo que os trabalhadores menos qualificados tendem a ser mais afetados.

É praticamente consenso a necessidade de requalificar os profissionais investindo em educação. No Brasil as barreiras são tremendas: a) de acordo com o Indicador de Analfabetismo Funcional (INAF), três a cada dez brasileiros têm limitação para ler, interpretar textos, identificar ironia e fazer operações matemáticas, sendo considerados analfabetos funcionais, contingente que representa 30% da população entre 15 e 64 anos; e b) o baixo desempenho no Programa Internacional de Avaliação de Estudantes (PISA) da Organização para Cooperação e Desenvolvimento Econômico (OCDE) sugere, além da baixa qualidade do ensino, uma grande disparidade nos resultados dependendo do contexto socioeconômico dos alunos.

Ensino deficiente é uma barreira aos trabalhadores no uso de tecnologias digitais, gerando, portanto, uma desigualdade digital de segundo nível que restringe ainda mais a mobilidade dos trabalhadores, particularmente os de baixa qualificação. Outro agravante é que o aprendizado formal, mesmo supondo qualidade para todos, por si só não preenche as lacunas de habilidades associadas às novas funções. As *soft skills* derivam de formação (não de treinamento) e são adquiridas em múltiplas vivências e experiências, em geral, acessíveis apenas a um contingente restrito da população.

Visando à sustentabilidade do trabalho, as políticas públicas em parceria com o setor privado precisam incorporar os conceitos de *lifelong learning* (aprendizagem ao longo da vida), fundamental para acompanhar a aceleração atual; e de autorregulação da aprendizagem, em que cada indivíduo estrutura, monitora e avalia seu próprio aprendizado, ampliando sua capacidade de retenção de conteúdo e engajamento. O futuro deverá privilegiar quem “aprende a aprender”.



Justiça e ética na IA

O aumento do agenciamento da IA, especialmente quando substitui a agência humana, remete a uma questão ética cada vez mais urgente: a responsabilidade. A sociedade humana, desde Aristóteles, considera que cada um é responsável pelas consequências de seus atos. Em *Ética a Nicômaco*, o filósofo acrescenta uma condição à responsabilidade moral, que é estar ciente do que se está fazendo, o que não é o caso dos algoritmos de IA: como agente, esses algoritmos são capazes de realizar ações e tomar decisões com consequências éticas, mas não são capazes de um pensamento moral, portanto, não podem ser responsabilizados pelas consequências de seus atos. Ou seja, os algoritmos de IA são agentes, mas não são agentes morais, porque, entre outros, carecem de consciência, emoções e sentimentos, intencionalidade. Segundo Aristóteles, somente os seres humanos realizam atos voluntários, cabendo aos humanos a responsabilidade pelos atos maquínicos. Ao delegar o agenciamento à IA, os seres humanos retêm a responsabilidade.⁵³

Atribuir aos humanos a responsabilidade sobre as decisões automatizadas, ou mesmo sobre meras previsões, não é trivial. Os sistemas de IA mais complexos, portanto, com maior probabilidade de causar danos, agregam contribuições de diversos desenvolvedores, utilizam bases de dados originadas de múltiplas fontes para treinar seus algoritmos e, por fim, são aplicados para executar tarefas, às vezes, em distintos domínios. A responsabilidade, desse modo, teria de ser distribuída entre todas as partes envolvidas (o que nem sempre é fácil de rastrear).

Assumir a responsabilidade sobre algo significa ser capaz de explicar o fenômeno e/ou evento. No caso do médico, por exemplo, é legítimo que o paciente demande explicação sobre os procedimentos recomendados, a mesma lógica aplica-se ao juiz, ao professor, ao profissional de RH, em suma, a todos os agentes com prerrogativas de tomar decisões com impactos na vida de terceiros. A responsabilidade e, consequentemente, a explicabilidade são um dos temas éticos mais sensíveis: por um lado, temos a opacidade intrínseca à técnica de redes neurais profundas (como os algoritmos correlacionam os parâmetros contidos nos dados), o que limita a capacidade dos humanos de explicar plenamente a decisão do sistema; por outro, dado o estágio de desenvolvimento da técnica, seus resultados refletem, em grande parte, decisões humanas, que, como comentado anteriormente, são fragmentadas, dificultando uma explicação abrangente. Ademais, grande parte dos usuários da IA carecem de conhecimento básico, logo, não estão aptos a oferecer nenhuma explicação.

Este bloco contém oito artigos sobre temas diversos, tais como *fake news* nos negócios, ética como objeto da ação humana e ameaça das tecnologias de reconhecimento facial. Inclui temas específicos, como o *IA Guidebook*, do Google, e o documentário *Coded Bias*.

Dois professores da Universidade de Washington, Jevin West e Carl Bergstrom, criaram o jogo *online Which Face Is Real* (*Qual rosto é real*) com base em milhares de rostos humanos virtuais artificiais desenvolvidos pela dupla. O desafio consiste em adivinhar qual rosto é verdadeiramente humano. Meio milhão de jogadores disputaram seis milhões de rodadas. A tecnologia do jogo é da Nvidia, empresa de processadores gráficos, e usa redes neurais (*deep learning*/ inteligência artificial) treinadas num imenso conjunto de retratos de pessoas. O percentual de acertos girou em torno de 60% na primeira tentativa, atingindo 75% de precisão em tentativas posteriores. Segundo seus criadores, a intenção foi alertar a sociedade sobre a capacidade tecnológica atual de gerar imagens falsas, e o risco é a impossibilidade de evitar usos não nobres dessa tecnologia.

Em outro exercício acadêmico, dois pesquisadores da Global Pulse, iniciativa ligada à Organização das Nações Unidas (onu), usando apenas recursos e dados de código-fonte aberto, mostraram com que rapidez poderiam colocar em funcionamento um falso gerador de discursos de líderes políticos em assembleias da onu. O modelo foi treinado em discursos proferidos por líderes políticos na Assembleia Geral da onu entre os anos 1970 e 2015. Em apenas 13 horas e a um custo de 7,80 dólares (despesa com recursos de computação em nuvem), os pesquisadores conseguiram produzir discursos realistas sobre uma ampla variedade de temas sensíveis e de alto risco, de desarmamento nuclear a refugiados.

O tema das *fake news* ganhou visibilidade pelos impactos negativos nos processos eleitorais, sobretudo na eleição de Donald Trump, em 2016, com os *bots* russos se passando por eleitores norte-americanos; no Brasil, a eleição de 2018 disseminou o uso de robôs e tecnologias de impulsionamento automático de mensagens, visando influenciar os eleitores. A produção de conteúdo falso está não só proliferando, como também se sofisticando: agregando inteligência artificial, despontam as *deep fakes*.

O fenômeno de falsificação na internet extrapola o âmbito das notícias e da política, atingindo igualmente o mundo dos negócios, particularmente as plataformas centradas em dados. O Review Meta, site independente que monitora a veracidade do *feedback online* da Amazon, identificou um crescimento de avaliações na plataforma postadas por usuários que não compraram o item em questão, ou seja, não o experimentaram. Não por coincidência, 98,2% dessas postagens avaliam o produto em cinco estrelas. A socióloga turca Zeynep Tufekci, em artigo na revista *Wired*, alerta que as alegações de falsidade também podem

ser falsas: “Na Amazon, você dificilmente pode comprar um filtro solar simples sem encontrar avaliações que alegam que o produto é falsificado. Aliviado por ter sido avisado, você pode ficar tentado a clicar. Mas talvez essa revisão em si seja falsa, plantada por um concorrente”.⁵⁴

O modelo de negócio do Google e do Facebook, para citar dois dos gigantes de tecnologia, baseia-se em oferecer aos anunciantes acesso segmentado aos potenciais consumidores, tornando mais assertivas as campanhas publicitárias *online*. Observa-se, contudo, que esse modelo também está suscetível a fraudes, repleto de visualizações e cliques falsos. Em 2016, o Facebook admitiu ter exagerado na quantificação do tempo que seus usuários assistem a vídeos na plataforma, caracterizando como um “erro” com efeito zero sobre o faturamento. Aparentemente, não foi esse o entendimento de muitos pequenos anunciantes: em 2018 entraram com uma ação coletiva alegando que a rede social estava inflando seus números propositalmente.

São muitos os exemplos mundo afora. Na Bulgária, em 2017, por exemplo, o Spotify foi fraudado no valor de 1 milhão de dólares: foram geradas músicas de 30 segundos (tempo médio de escuta), e contas falsas automatizadas para reproduzi-las; os fraudadores embolsavam a diferença entre os *royalties* e a quantia paga à plataforma para listar suas próprias faixas.

Vivemos um período de crise generalizada de confiança, que extrapola os eventos na internet. Acima de regras morais e éticas, arcabouço regulatório e sistemas de punição, a sociedade precisa de um mínimo de confiança entre seus agentes – instituições, governos e cidadãos – para funcionar de maneira sadia. As facilidades da tecnologia e do meio digital só exacerbam o ambiente atual da sociedade.

Os algoritmos de inteligência artificial podem ser éticos?

9.8.2019

Em seu último livro, *Máquinas como eu*, o escritor inglês Ian McEwan trata da distinção moral entre Miranda, uma jovem de 22 anos vizinha-namorada de Charlie, e Adão, o humanoide adquirido por ele com recursos herdados pela morte de sua mãe.⁵⁵ O autor atribui ao humanoide uma visão moral mais consistente e, indo além, levanta a possibilidade de nós, seres humanos, sermos capazes de criar seres artificiais moralmente superiores (suposição ficcional, no momento não existe nenhuma base científica para tal).

O tema da ética permeia a sociedade humana desde Aristóteles e foi mudando de sentido ao longo da história, resguardando, contudo, a crença de que apenas o humano é dotado da capacidade de pensar criticamente sobre valores morais e dirigir suas ações em termos de tais valores. Com o avanço recente das tecnologias de inteligência artificial, as questões éticas estão na pauta. Associados à robótica, como no caso do humanoide Adão de McEwan, ou mediando as interações sociais e os processos decisórios, os algoritmos de inteligência artificial agregam inúmeros benefícios, mas, simultaneamente, carecem de transparência, são difíceis de ser explicados e comprometem a privacidade. Diariamente, aparecem casos ilustrativos no Brasil e mundo afora.

Nos estados de Utah e Vermont, nos Estados Unidos, o Departamento Federal de Investigação (fbi) e o Serviço de Imigração e Alfândega (ice) usaram tecnologia de reconhecimento facial na análise de milhões de fotos de carteiras de habilitação com o propósito de identificar imigrantes ilegais. A questão ética nesse procedimento é que, aparentemente, não houve conhecimento, muito menos consentimento, dos motoristas; ademais, vários estudos indicam que os modelos de reconhecimento de imagem não são perfeitos, em alguns casos a margem de erro pode ser relevante, em função, entre outros, do viés contido nos dados.

Em 2018, o Facebook lançou um aplicativo que estimulava seus usuários a postarem fotos atuais e de 10 anos atrás. Em julho de 2019, o FaceApp alcançou o primeiro lugar na lista geral de aplicativos do Google Play e da App Store, envelhecendo as fotos e projetando aparência futura. Ambos foram sucesso e viralizaram. Longe de serem um mero entretenimento, esses aplicativos servem para captar dados e utilizá-los no treinamento dos algoritmos de reconhecimento de imagem (inteligência artificial/*deep learning*). Em ambos os casos houve consentimento dos usuários, que aderiram voluntariamente ao desafio, mas não houve transparência quanto ao propósito.

Em meados de 2017, pesquisadores da Universidade de Stanford tornaram

público um algoritmo de inteligência artificial, o Gaydar, com a finalidade de, com base nas fotografias das plataformas de namoro, identificar os homossexuais. A motivação inicial era protegê-los, contudo, a iniciativa foi vista como potencial ameaça à privacidade e à segurança, desencadeando inúmeros protestos.

Nos Estados Unidos – país com, provavelmente, o mais eficiente arcabouço legal de proteção aos seus cidadãos e instituições –, existe o Institutional Review Board (Conselho de Avaliação Institucional, irb), que é um comitê independente voltado para garantir a ética nas pesquisas e que norteia os conselhos dos centros de pesquisa e universidades; o estudo que originou o Gaydar foi previamente aprovado pelo Conselho de Avaliação de Stanford. A questão é que as regras foram fixadas há 40 anos. “A grande e vasta maioria do que chamamos de pesquisa de ‘grandes dados’ não é abrangida pela regulamentação federal”,⁵⁶ diz Jake Metcalfe, do Data & Society, instituto de Nova York dedicado aos impactos sociais e culturais do desenvolvimento tecnológico centrado em dados.

No evento Sustainable Brands em São Paulo, David O’Keefe, da Telefonica Dynamic Insights, controladora da operadora de telefonia móvel Vivo, apresentou alguns produtos derivados dos dados captados das linhas móveis (*mobile phone data*). Com o título “Using Global Comms Data and Machine Learning to Provide Digital Relationship Insights in Multinationals” (Usando dados comuns globais e aprendizado de máquina para fornecer informações de relacionamento digital em multinacionais), O’Keefe descreveu o “produto” em que, por meio dos dados dos celulares dos funcionários de uma empresa multinacional (quem ligou para quem, com que frequência, quanto tempo durou a ligação etc.), é possível identificar as redes informais internas, importante elemento nas estratégias de gestão (sem conhecimento e consentimento dos usuários).

Se no Rio de Janeiro e em outros estados os órgãos de segurança estão usando livremente a tecnologia de reconhecimento facial, em São Francisco, em 14 de maio último, foi proibido o uso pela polícia e por outros órgãos da administração municipal. São Francisco é a primeira grande cidade dos Estados Unidos a proibir o uso da tecnologia de reconhecimento facial como aparato de vigilância e controle público.

Os modelos estatísticos buscam padrões e fazem previsões, contudo, seus resultados não são objetivos nem garantidos, em parte, porque são baseados em amostras que nem sempre são representativas do universo total (incerteza, margem de erro). Adicionalmente, os fatores intangíveis não são quantificáveis.

Se em muitas situações do cotidiano a imprecisão não incomoda, o mesmo não se pode dizer, por exemplo, de processos relacionados à saúde; até se aceita que os algoritmos de ia diagnostiquem tumor cancerígeno, mas dificilmente o paciente aceita em uma quimioterapia automatizar a decisão do tipo de medicação e da dose.

Uma das críticas, legítima, é que esses sistemas são caixas-pretas – não são

transparentes, explicáveis –, mas devemos lembrar que os humanos nem sempre sabem explicar o porquê de determinadas decisões; a diferença, talvez, é que os humanos, como seres racionais, inventam explicações, produzem justificativas aparentemente plausíveis, mas nem sempre fidedignas, o que as máquinas não são capazes de fazer.

O avanço recente da inteligência artificial, quando as máquinas não seguem mais processos de decisão pré-programados pelos humanos e começam a “aprender” por si mesmas (*machine learning, deep learning*), coloca para a sociedade novos desafios éticos e a premência de estabelecer arcabouços legais a partir de uma regulamentação que, simultaneamente, proteja os indivíduos e as instituições, e preserve o grau de liberdade necessário ao desenvolvimento científico e comercial. Será que a lei brasileira de proteção de dados dá conta dessa complexidade?

Alerta: as tecnologias de reconhecimento facial estão nos ameaçando

4.10.2019

No Carnaval de 2019, fato amplamente divulgado na mídia, o sistema de reconhecimento facial da polícia baiana localizou e prendeu um criminoso fantasiado de mulher no circuito Dodô. A polícia do Rio de Janeiro, em 2018, contratou o sistema britânico Facewatch com o propósito de identificar 1.100 criminosos ao cruzarem as câmeras de segurança; segundo anunciado, o sistema vem sendo utilizado no Reino Unido há cerca de sete anos e conta com 30 mil câmeras espalhadas pelo país.

Parte dos sistemas de vigilância usam tecnologias de reconhecimento facial com inteligência artificial, especificamente a técnica de *deep learning*, cujos resultados, como todo modelo estatístico de probabilidade, não são precisos. Nos Estados Unidos, o Instituto Nacional de Padrões e Tecnologia (nist) indicou que entre 2014 e 2018 a precisão passou de 96% para 99,8%, mas em condições especiais testadas em laboratórios.

Inúmeros casos têm sido relatados globalmente sobre erro de identificação, alguns com danos relevantes, como o ocorrido em julho último, no Rio de Janeiro: uma mulher foi detida por engano em Copacabana e levada à delegacia do bairro, após as câmeras de reconhecimento facial darem positivo. Em paralelo, como não poderia deixar de ser, surgem soluções para “enganar” os sistemas de reconhecimento facial, denominadas *deep learning adversarial*, que provocam uma “ilusão de ótica” nas máquinas.

Proliferam câmeras de vigilância, em espaços públicos e privados, sem a necessária consciência da sociedade sobre os riscos. Elas captam nossa imagem nos aeroportos – o Departamento de Segurança Interna dos Estados Unidos estima que o reconhecimento facial examinará 97% dos passageiros de companhias aéreas até 2023 –, metrô, espaços comerciais, dispositivos digitais. Estão implantadas, ou em vias de, em escolas, transportes públicos, locais de trabalho, unidades de saúde e nas ruas de algumas cidades e regiões.

A potencial ameaça à privacidade tem suscitado fortes reações contrárias a esses sistemas. Em 14 de maio deste ano, São Francisco tornou-se a primeira cidade dos Estados Unidos a proibir o uso dessa tecnologia pela polícia e por outros órgãos da administração. A medida, deliberada pelo Conselho de Supervisores da cidade em votação de 8x1, determina igualmente que todos os departamentos divulguem as tecnologias de vigilância utilizadas ou em processo de desenvolvimento, e definam políticas a esse respeito a serem aprovadas pelo Conselho. Em julho último, o conselho municipal de Oakland, também na

Califórnia, votou a favor de uma lei que proíbe o uso de tecnologias de reconhecimento facial pelas agências públicas, tornando-se, após Somerville, em Massachusetts, a terceira cidade norte-americana a aprovar leis semelhantes em 2019.

Em setembro, o Ada Lovelace Institute, órgão inglês de pesquisa independente, divulgou os resultados de uma enquete realizada no Reino Unido sobre a percepção do público com relação ao uso de tecnologia de reconhecimento facial, indicando que: a) 90% dos ingleses estão conscientes do uso da tecnologia, mas apenas 53% admitem conhecer do que se trata; b) 46% da população acredita que seja seu direito consentir ou não sobre a captação e o uso de sua imagem, percentual que sobe para 56% entre as minorias étnicas; c) os ingleses, em geral, estão dispostos a aceitar o reconhecimento facial em larga escala se houver um benefício explícito, e 70% acham que deva ser permitido para uso da polícia em investigações criminais; d) 67% não se sentem confortáveis com o seu uso nas escolas, e 61% não se sentem confortáveis com o seu uso nos transportes públicos; e) os ingleses esperam regulamentação governamental, salvaguardas e limitações no uso policial, e 55% defendem que o uso da tecnologia seja limitado a circunstâncias específicas; e f) os ingleses, majoritariamente, não confiam no setor privado no uso ético do reconhecimento facial, 77% não se sentem confortáveis com o uso em lojas para rastrear clientes, e 76%, por departamentos de rh no recrutamento de candidatos para empregos de nível inicial. Cresce o debate público na Inglaterra sobre aspectos éticos dessas tecnologias, acompanhado de protestos, críticas políticas e processos legais. Existe um clamor social contra a ausência de consultas públicas adequadas.

Em agosto de 2020, entra em vigor no Brasil a Lei Geral de Proteção de Dados (lgpd), que, se ainda não está comprovada sua capacidade de fiscalização, já representa um avanço pelo simples fato de existir um arcabouço legal sobre o uso de dados pessoais. Antecipando-se aos termos da lei, em fevereiro último, o Instituto Brasileiro de Defesa do Consumidor (Idec) notificou a varejista de roupas Hering para que explicasse a finalidade dos dados captados em sua loja do Morumbi Shopping, em São Paulo: câmeras com tecnologia de reconhecimento facial captam as reações dos clientes às peças expostas nas araras, e, em paralelo, sensores identificam suas preferências ao circularem pela loja. O mesmo instituto, em 2018, processou a ViaQuatro, concessionária da linha 4-Amarela do metrô de São Paulo, pelo uso de sensores com tecnologia de reconhecimento facial nos seus painéis publicitários; a Justiça determinou a suspensão imediata do uso dessa tecnologia, alegando falta de transparência sobre a finalidade, o tratamento e o uso das imagens.

Os modelos de negócio emergentes são, em parte, baseados em dados, em captar e extrair informações valiosas dos dados, o que gera uma inédita relação entre ética e negócios, deixando de ser associada exclusivamente à reputação de marca e passando a fazer parte dos resultados financeiros e comerciais.

A objetividade relativa da ia pode neutralizar a subjetividade humana?

27.12.2019

Cotidianamente, na vida profissional e pessoal, tomamos decisões. Ainda iludidos pelas ideias do Iluminismo do século XVIII – razão, livre-arbítrio –, acreditamos que esses processos cognitivos sejam controlados, isto é, que sabemos as razões pelas quais estamos decidindo. Quando confrontados, em geral, os seres humanos explicam suas decisões, mas essas explicações são fabricações racionais. A verdade é que desconhecemos o que de fato influenciou nossa decisão, motivações em parte subjetivas e inconscientes. Sem contar a falta de transparência deliberada – sabemos os reais critérios de decisão na seleção para uma vaga de emprego? As organizações, em diversas situações, justificam a opacidade de seus processos com o argumento da proteção da propriedade intelectual e do sigilo comercial.

Os modelos de decisão automatizados trazem inúmeros benefícios aos indivíduos e à sociedade, contudo, contêm riscos. Entre outros, perpetuam os preconceitos e geram assimetria de informação entre os indivíduos e as instituições detentoras de grandes conjuntos de dados (cujos modelos de negócio são baseados em extrair informações úteis desses dados). O fato é que não sabemos por que os modelos de inteligência artificial fazem as escolhas que fazem, e essa dificuldade cresce proporcionalmente ao aumento da complexidade dos próprios modelos.

Esforços estão sendo canalizados para garantir equidade, evitando que as decisões sejam influenciadas por características de gênero, raça ou qualquer outro atributo exclusivo de um grupo, cuidando para que as amostras usadas no treinamento dos algoritmos reflitam a integralidade da população; assim como para diversificar as equipes de desenvolvedores, responsáveis por introduzir e enfatizar nos modelos determinadas variáveis em detrimento de outras.

A transparência do como e do por que as decisões automatizadas foram geradas permitirá que os afetados compreendam os motivos e fundamentem suas dúvidas e questionamentos. Na saúde, por exemplo, essa falta de transparência é um obstáculo tanto para os médicos quanto para os pacientes, que não se sentem confortáveis com diagnósticos automatizados.

Não por acaso, em paralelo à proliferação de decisões automatizadas em distintas esferas – filtram, classificam, recomendam, personalizam e modelam a experiência humana desde diagnósticos médicos até processos jurídicos –, surgem organizações dedicadas a combater, denunciar, fiscalizar e desenvolver soluções que respondam à pergunta: “Como funcionam esses modelos?”.

Em março de 2019, o Instituto de Engenheiros Eletricistas e Eletrônicos (Institute of Electrical and Electronics Engineers - IEEE) dos Estados Unidos – a maior organização profissional sem fins lucrativos dedicada ao avanço da tecnologia em benefício da humanidade – divulgou um relatório com foco na ética, visando informar, aprimorar e apoiar as organizações dedicadas à inteligência artificial. O documento contém princípios éticos a serem incorporados aos modelos de decisão, a partir de três macroprincípios: inteligibilidade, processo técnico transparente e explicável; precisão, que os resultados (ou *outputs*) representem a verdade; e equidade, que os dados e os algoritmos contemplem uma amostra representativa do universo em questão.

O Fórum Econômico Mundial identificou quase 300 esforços mundo afora com foco em desenvolver princípios éticos para a inteligência artificial, envolvendo órgãos governamentais, universidades e associações profissionais como a Association for the Advancement of Artificial Intelligence (AAAI). O G7, grupo dos países mais industrializados, propôs um Painel Internacional de Inteligência Artificial inspirado no Painel Intergovernamental sobre Mudanças Climáticas da ONU.

A tão criticada opacidade deve ser enfrentada. Trata-se de uma oportunidade de gerar decisões transparentes, o que nunca teremos com os seres humanos – as pessoas não são isentas nem neutras. Existem indícios científicos de que será possível mitigar, e até eliminar, o famoso viés nos algoritmos inteligentes. Parte dos avanços são compartilhados na ACM Conference on Fairness, Accountability, and Transparency (FACAT), conferência interdisciplinar anual dedicada a investigar e resolver o problema do viés nos sistemas inteligentes.

A objetividade relativa da inteligência artificial pode vir a neutralizar a subjetividade humana. Essa é uma de suas promessas. Até lá, regulamentações adequadas, como as leis de proteção de dados, podem minimizar seus efeitos discriminatórios.

Aplicativo de reconhecimento facial: como fica o direito ao anonimato?

24.1.2020

A ClearView, empresa criada por Hoan Ton-That, programador australiano e ex-modelo, e Richard Schwartz, assessor de Rudolph Giuliani quando prefeito de Nova York, é um aplicativo de reconhecimento facial que ameaça o direito do cidadão de andar anonimamente pelas ruas da cidade. Com um banco de dados de mais de três bilhões de imagens extraídas de sites na internet, o aplicativo é capaz de identificar fotos públicas de qualquer pessoa, ou seja, sua localização em qualquer lugar público. O aplicativo pode localizar, igualmente, um militante em uma manifestação ou um criminoso, revelando informações privadas como nome e endereço de moradia.

Dos dois primeiros engenheiros contratados pelo aplicativo, em 2016, um deles desenvolveu um programa de coleta de imagens faciais na internet, acessando sites de emprego, de notícias, educacionais e redes sociais, incluindo Facebook, YouTube, Twitter e Instagram. O segundo engenheiro aperfeiçoou um algoritmo de reconhecimento facial com base em inteligência artificial (rede neural/*deep learning*), capaz de converter as imagens em fórmulas matemáticas a partir da geometria facial (distância entre os olhos, por exemplo). Ao carregar uma foto no aplicativo, automaticamente aparecem todas as cópias dessa foto armazenadas no banco de imagens e os links dos sites que publicaram a foto original.

Além de não ser do conhecimento público, a captação de imagens pela Clearview é ilegal e viola os termos de serviço dos sites: as plataformas de mídia social, incluindo o Facebook, proíbem a captura de imagens de seus usuários por terceiros. Ton-That minimiza a infração, alegando que essa é uma prática comum e que, além do mais, o aplicativo usa apenas imagens publicamente disponíveis (não tem acesso, por exemplo, àquelas que são protegidas pela configuração de privacidade do Facebook). Contudo, Jay Nancarrow, porta-voz do Facebook, informou que o aplicativo está sendo analisado e que serão tomadas medidas apropriadas no caso de comprovação de violação das regras.

Em matéria publicada no *The New York Times*, a jornalista Kashmir Hill denuncia as atividades do aplicativo Clearview e suas conexões com departamentos de polícia norte-americanos e o fbi (total de prestação de serviço: mais de 600 agências policiais desde 2019).⁵⁷ Segundo Hill: “A técnica de vendas mais eficaz da empresa era oferecer testes gratuitos de 30 dias aos policiais, que incentivaram seus departamentos a se inscreverem e recomendaram o aplicativo para oficiais de outros departamentos de polícia. De acordo com a própria

empresa e documentos fornecidos pelos departamentos de polícia, o Sr. Ton-That finalmente teve seu sucesso viral”.

Quando a jornalista começou a pesquisar a Clearview, em novembro de 2019, o site da empresa era uma página vazia com um endereço falso de Manhattan. No perfil da empresa no LinkedIn constava um gerente de vendas, John Good, que na verdade era o próprio fundador, Ton-That. “Durante um mês, as pessoas afiliadas à empresa não retornaram meus e-mails ou telefonemas. Enquanto a empresa estava se esquivando, também estava me monitorando. A meu pedido, vários policiais passaram minha foto pelo aplicativo Clearview, em seguida receberam telefonemas de representantes da empresa perguntando se estavam conversando com a mídia – um sinal de que a Clearview tem a capacidade e, nesse caso, o apetite de monitorar quem a está procurando”, conta Hill.

Além das imagens captadas ilegalmente das redes sociais, as agências policiais estão fazendo *upload* de fotos confidenciais nos servidores da Clearview sem nenhuma garantia sobre sua competência em proteger os dados e a privacidade dos cidadãos. Aparentemente, os policiais estão entusiasmados com o aplicativo pela ajuda efetiva que receberam na identificação de vários criminosos em diversos estados norte-americanos.

A polícia estadual de Indiana foi o primeiro cliente da Clearview. Em fevereiro de 2019, um crime foi resolvido em 20 minutos com o auxílio do aplicativo: dois homens entraram em uma briga num parque, tendo um atirado no estômago do outro; um cidadão registrou o crime no celular, fornecendo à polícia o rosto do atirador; o aplicativo localizou o criminoso em um vídeo legendado com o seu nome postado nas mídias sociais; em seguida, ele foi preso e acusado.

A invasão da privacidade é o efeito perverso mais evidente, comprometendo a liberdade de andar pelas ruas das cidades sem ser reconhecido, mas não é o único. Como todos os modelos estatísticos baseados em redes neurais, seus resultados indicam a probabilidade de algo ocorrer em percentuais menores do que 100%; no caso da Clearview, seu fundador admite que o aplicativo tem limitações em decorrência, principalmente, das imperfeições das fotos, atingindo assertividade de 75% (vale ressaltar que o aplicativo não foi testado por nenhum órgão especializado, como o Instituto Nacional de Padrões e Tecnologia, agência federal norte-americana que avalia o desempenho de algoritmos de reconhecimento facial). O percentual de 25% de erro pode parecer pouco, mas com certeza não foi para a mulher detida por engano em Copacabana, em julho de 2019, após as câmaras de reconhecimento facial da polícia do Rio de Janeiro a identificarem erroneamente.

“Procurar alguém pelo rosto pode se tornar tão fácil quanto pesquisar um nome no Google. Estranhos seriam capazes de ouvir conversas sensíveis, tirar fotos dos participantes e conhecer segredos pessoais. Alguém andando na rua

seria imediatamente identificável – e seu endereço residencial estaria a apenas alguns cliques de distância. Anunciaria o fim do anonimato público”, sentencia Kashmir Hill.

Ética é objeto da ação humana, não existe ética da inteligência artificial

17.7.2020

A ética é objeto da ação humana, as tecnologias não têm objetivos próprios nem autonomia. No caso da inteligência artificial, em seu estágio atual de desenvolvimento, em que o humano detém a prerrogativa de controle, não há como conceder a esses sistemas o status moral. A ia não tem uma ética própria, trata-se de elaborar um conjunto de melhores práticas que possa ser replicado em uma ampla variedade de configurações. O que não é nada trivial, dada a complexidade de seus sistemas.

Jess Whittlestone, pesquisadora sênior do Leverhulme Center for the Future of Intelligence, da Universidade de Cambridge, em um artigo com mais três colegas publicado na *Nature Machine Intelligence*, alerta sobre a urgência de encontrar maneiras de incorporar a ética no desenvolvimento e na aplicação da inteligência artificial, e não como uma reflexão tardia. Para os autores, nos últimos anos a ética em ia se concentrou em princípios gerais que não dizem nada sobre como proceder quando esses princípios entram em conflito, por exemplo, no combate à covid-19 – como equilibrar o potencial da ia para salvar vidas com as ameaças aos direitos civis, como privacidade? Além disso, esses princípios éticos gerais têm pouco efeito sobre o desenvolvimento e a aplicação das tecnologias.⁵⁸

A Comissão Europeia lançou, em 8 de abril de 2019, o guia *Ethics Guidelines for Trustworthy AI (Orientações Éticas para uma ia Confiável)*,⁵⁹ cujos princípios têm sido replicados mundo afora. Sua elaboração contou com 52 especialistas, incluindo membros de organizações não governamentais, acadêmicos e representantes de empresas de tecnologia. Em fase-piloto ao longo de 2020, o guia está aberto ao debate e às contribuições da sociedade (consulta pública). As principais diretrizes do documento são o respeito à autonomia humana, a prevenção contra o dano ao ser humano, a explicabilidade e a transparência, e a justiça (evitar trajetórias oblíquas que levem à discriminação). Além disso, propõe prerequisites para uma ia confiável, tais como intervenção e supervisão humana, robustez técnica e segurança, privacidade e governança de dados, bem-estar social e ambiental, e prestação de contas.

Esses princípios gerais estão na base fundadora de diversos institutos, tais como o Future of Life Institute, constituído em 2014 pelo professor do mit Max Tegmark, com financiamento de Elon Musk; o ai Now Institute, da Universidade de Nova York, fundado em 2017; e o ai for Good Institute, da Universidade de Stanford, de 2019. O Future of Humanity Institute, liderado pelo filósofo inglês Nick Bostrom, criado em 2005, tem hoje um centro de governança em

inteligência artificial.

As gigantes da tecnologia, pressionadas pela sociedade, também produzem suas listas de princípios. O Google, por exemplo, em 2018 anunciou sete objetivos essenciais para criar sistemas socialmente benéficos, seguros e imparciais, com maior responsabilidade. Em 2019, divulgou o lançamento de um conselho consultivo externo de ética em inteligência artificial (alguns funcionários reagiram questionando sua legitimidade). Um ano antes, sua subsidiária DeepMind criou o grupo de estudo DeepMinds Ethics & Society (dmes), dedicado ao estudo dos impactos da ia na sociedade, liderado por Verity Harding e Sean Legassick e mais 25 pesquisadores com dedicação exclusiva, com colaboração de Nick Bostrom e parcerias com o ai Now Institute e o Leverhulme Center for the Future of Intelligence.

Essas iniciativas, aparentemente, não norteiam a condução de seus negócios: em fevereiro último, por exemplo, o Google foi denunciado pela Comissão Irlandesa de Proteção de Dados (dpc), local de sua sede na Europa, sobre o tratamento de dados de geolocalização de seus usuários. O órgão regulador alegou ter recebido inúmeras queixas de associações europeias de consumidores de que o Google disponibilizou dados pessoais de seus usuários à anunciantes, em desacordo com o Regulamento Geral sobre a Proteção de Dados da Europa (gdpr), além de obter vantagens sobre modelos concorrenciais semelhantes ao gerar segmentação mais precisa, consequentemente, atraindo mais anunciantes.

O ponto comum entre esses movimentos é que propõem princípios generalistas, de aplicabilidade restrita, de difícil tradução em boas práticas para nortear o ecossistema de ia (pesquisadores, desenvolvedores, instituições, universidades, empresas, governos). Além disso, parte desses princípios gerais não leva em conta as limitações atuais da tecnologia que impossibilitam, por exemplo, a eliminação do viés e a superação da opacidade (não explicabilidade). Outro complicador é que não basta aplicar quaisquer desses princípios éticos apenas no estágio de desenvolvimento e implementação da tecnologia: os sistemas de ia “aprendem” e evoluem continuamente; um sistema alinhado na partida, com novos conjuntos de dados e novo aprendizado, pode sair do alinhamento, o que demandaria monitoramento contínuo.

O Berkman Klein Center for Internet & Society, da Universidade de Direito de Harvard, parece estar mais próximo de enfrentar o hiato entre os princípios éticos gerais e um arcabouço ético aplicado ao desenvolvimento de tecnologias de inteligência artificial. No documento “Ethics and Governance of ai at Berkman Klein: Report on Impact, 2017-2019”, publicado em outubro de 2019, é possível compreender a dimensão e a diversidade da atuação do instituto na comunidade da Universidade Harvard, na comunidade de direito do Estado, nas empresas e no setor público.⁶⁰

O ponto de partida foi a agregação em oito temas dos cerca de 50 princípios globais de inteligência artificial: privacidade, responsabilidade, segurança e

proteção, transparência e explicabilidade, justiça e não discriminação, controle humano da tecnologia, responsabilidade profissional e promoção dos valores humanos. A partir desse guia geral, o instituto se organizou em “trilhas”. Por exemplo, o agtech Forum interage com os advogados e suas equipes, visando acelerar as questões relacionadas à privacidade e à segurança cibernética.⁶¹

O programa Assembly reúne anualmente a indústria, a academia e o governo para trabalhar em projetos de tecnologia direcionados ao bem social, com resultados concretos nos últimos anos, como o Data Nutrition Project, agora uma entidade sem fins lucrativos, que elaborou novos padrões e formatos – inspirados nos rótulos nutricionais do fda para produtos alimentícios – para avaliar e rotular conjuntos de dados.

Outra de suas iniciativas é o projeto Techtopia, que reúne equipes de professores e estudantes com foco no desenvolvimento de novas pedagogias. Em maio de 2019, a equipe de jovens do instituto lançou o relatório intitulado Youth and Artificial Intelligence: Where We Stand (Juventude e inteligência artificial: onde estamos).⁶²

O *workshop* Towards Global Guidance on ai and Child Rights (Orientação global sobre ia e direitos da criança), promovido pelo Fundo das Nações Unidas para a Infância (unicef), em junho de 2019, examinou algumas das maneiras como a tecnologia de inteligência artificial está reconfigurando as experiências dos alunos desde a primeira infância e sinalizou uma série de perguntas-chave para professores e formuladores de políticas.⁶³ O Principled Artificial Intelligence Project, lançado na Cyberlaw Clinic, em junho de 2019, trabalha para mapear o conjunto de diretrizes, padrões e práticas recomendadas sobre desenvolvimento e implantação de ia. O projeto metalab ai + Art, da Universidade Harvard, produziu, durante dois anos (de maio de 2017 a maio de 2019), 10 projetos, realizou mais de 45 exposições em 11 países, foi coberto em mais de 25 artigos, ministrou nove oficinas e cursos e realizou mais de 50 palestras públicas.⁶⁴

Com a crescente presença da inteligência artificial em nossas vidas, é premente reduzir a lacuna de conhecimento entre os especialistas em ia e as pessoas que usam, interagem e são impactadas por essas tecnologias. É preciso, igualmente, ir além de simples “declaração de princípios”, e pensar em como reduzir as externalidades negativas de seus modelos e proteger a sociedade. No mínimo, espera-se que desenvolvedores, cientistas e empresas, no mínimo, explicitem os riscos associados aos seus modelos de IA, riscos conhecidos, ou seja, estamos falando de transparência.

Princípios éticos gerais *versus* regras práticas: reflexões sobre o *AI Guidebook* do Google

31.7.2020

O filósofo italiano Luciano Floridi, em artigo publicado em coautoria com o pesquisador Josh Cowls, reconhece que a inteligência artificial esteja se disseminando na sociedade com potencial de aliviar ou ampliar as desigualdades, e que os princípios éticos gerais propostos por instituições mundo afora são limitados como subsídios à criação de leis, regras, normas técnicas e melhores práticas.⁶⁵

Os autores propõem incluir um “novo” princípio: explicabilidade, “incorporando tanto o senso epistemológico de inteligibilidade (como resposta à pergunta ‘como isso funciona?’), quanto o sentido ético, de responsabilidade (como resposta à pergunta: ‘quem é responsável pela maneira como funciona?’)”. A proposta dos autores causa estranheza, primeiramente porque não se trata de um princípio novo, pelo contrário, está incluso na maioria das listas de princípios gerais; e, em segundo lugar, porque conflita com o funcionamento da própria tecnologia (a famosa “caixa-preta” ou não explicabilidade dos modelos de redes neurais/*deep learning*).

A atual relação entre tecnologia e seres humanos é complexa, como admite a Comissão Europeia, em documento de 2019, *High-Level Expert Group on Artificial Intelligence*: “Por um lado, as tecnologias são construídas por pessoas e organizações específicas, de modo que incorporam e reproduzem normas sociais, valores e outras forças econômicas, ecológicas, políticas e culturais; por outro lado, as tecnologias moldam como trabalhamos, nos movemos, nos comunicamos e vivemos”.⁶⁶ A Comissão alerta que as implicações éticas e sociais da tecnologia precisam considerar o entrelaçamento fundamental dos domínios humano e tecnológico – “os seres humanos são seres tecnológicos, assim como as tecnologias são entidades sociais” –, e ser capazes de distinguir os impactos no nível individual (autonomia, identidade, dignidade, privacidade e proteção de dados) e os impactos no nível social (justiça e equidade, identidade coletiva e boa vida, responsabilidade e transparência, democracia, confiança).

O ponto de partida desse documento, e também do documento *Recommendation of the Council on Artificial Intelligence*,⁶⁷ da Organização para a Cooperação e Desenvolvimento Econômico (ocde), foi o AI4People, primeiro fórum global da Europa sobre os impactos sociais da inteligência artificial, realizado em 10 de fevereiro de 2018, em Bruxelas. Reunindo mais de 50 especialistas independentes, pesquisadores, tomadores de decisão e representantes da indústria e da sociedade civil, o objetivo do fórum era esboçar um conjunto de

diretrizes éticas destinado a facilitar o desenho de políticas favoráveis ao desenvolvimento de uma “ia benéfica”.

Em paralelo com os organismos internacionais e institutos, o mercado se movimenta para adequar seus produtos e serviços às leis de proteção de dados, que indiretamente afetam os modelos de ia baseados em dados pessoais, e incorporar valores éticos e sociais aos seus sistemas inteligentes para atender à pressão da sociedade (preservar suas reputações, imagens de marca).

O People + ai Research (pair), área do Google criada em 2010, reúne uma equipe multidisciplinar coliderada pela brasileira Fernanda Viégas e por Martin Wattenberg: “Nosso objetivo é fazer pesquisas fundamentais, inventar novas tecnologias e criar estruturas para o design, a fim de conduzir uma abordagem centrada no ser humano à inteligência artificial”. Com base em recomendações, dados e *insights* de mais de 150 especialistas das equipes de produtos, com a colaboração de pesquisadores acadêmicos e consultores externos, em 2019, o pair lançou o *AI Guidebook*,⁶⁸ conjunto de diretrizes com foco em prover os designers e os gerentes de negócio de subsídios para “melhor representar o usuário na ‘mesa’ de desenvolvimento de produtos”, ou seja, defender o desejo do usuário por mais transparência e confiabilidade.

O guia é dividido em seis capítulos, e cada um deles contém uma planilha com orientações concretas, por exemplo, que os desenvolvedores, antes de iniciar o desenvolvimento de novos produtos, respondam a questões tais como: quem são seus usuários? Quais são seus valores? Qual problema você deve resolver para eles? Como você resolverá esse problema? Como você saberá quando a experiência for concluída? Há uma preocupação explícita em não decepcionar o usuário, nem superestimar o produto (atitude típica do mercado de tecnologia). A recomendação é dar ênfase aos benefícios, e não à tecnologia, sem, contudo, deixar de comunicar a natureza algorítmica e os limites dos produtos: sendo baseados em estatísticas e probabilidades, o usuário não deve confiar totalmente no sistema.

A parte referente à coleta e ao uso de dados (capítulo 3 do *AI Guidebook*) explicita um conjunto de regras de conduta visando adequar-se às leis de proteção de dados (privacidade, consentimento, segurança), minimizar o potencial viés das bases de dados e garantir o alinhamento às metas do produto e às necessidades do usuário.

Leitura atenta do guia indica, contudo, mais foco nos interesses comerciais (garantir satisfação do usuário na experiência com os produtos e serviços) e jurídico (proteção contra futuros passivos legais), e menos na transparência. Muitas ações poderiam ser tomadas em relação ao interesse do usuário; o tradutor do Google (*Google Translate*), por exemplo, poderia incorporar um aviso sobre suas limitações ou atribuir cores distintas ao grau de acurácia da tradução de cada palavra (essas ideias foram compartilhadas com Fernanda Viégas, a quem pareceram factíveis). O *AI Guidebook* ainda está em fase de implementação, é

cedo para verificar sua eficácia, mas pode servir de referência para o mercado, principalmente para as *startups*, geralmente sem times multidisciplinares.

Os modelos de inteligência artificial são complexos, dificultando a compreensão de usuários com conhecimento técnico limitado. É controversa, contudo, a crença de que a falta de transparência, a não explicabilidade, desses sistemas incomoda efetivamente a maioria dos usuários (aparentemente, contentam-se com os benefícios). De qualquer forma, certamente a sociedade precisa estar vigilante, e para tal é imprescindível conhecer a lógica e o funcionamento dessas tecnologias.

Documentário *Coded Bias*: o rosto como última fronteira da privacidade

16.4.2021

Coded Bias (2020), da cineasta e ativista norte-americana Shalini Kantayya, é o melhor documentário produzido sobre as externalidades negativas das tecnologias de inteligência artificial, pela qualidade e precisão das abordagens, pelo relato fidedigno dos fatos, pela seleção dos casos ilustrativos e pelos depoimentos de especialistas convidados. O foco são os sistemas de reconhecimento facial utilizados em larga escala mundo afora.

A protagonista do documentário é Joy Buolamwini, estudante de doutorado em Ciência da Computação do mit e fundadora da Algorithmic Justice League, cuja missão é defender e preservar os direitos e liberdades individuais garantidos pela Constituição dos Estados Unidos. O enredo começa com a descoberta de Buolamwini da discriminação de gênero contida num sistema de reconhecimento facial que seria usado num projeto de arte e tecnologia: o sistema só reagia quando Buolamwini, que é negra, colocava uma máscara branca. Impactada pela descoberta e motivada pelo livro de Cathy O’Neil *Algoritmos de destruição em massa*, Buolamwini empreendeu uma longa batalha contra o uso indiscriminado dessa tecnologia, substanciada em investigações científicas e análise de casos reais.⁶⁹ Em maio de 2019, por iniciativa dela e de seus colaboradores, ocorreu a primeira audiência no Congresso dos Estados Unidos sobre o uso de tecnologias de reconhecimento facial.

Um dos exemplos explorados no documentário é o uso de reconhecimento facial nos sistemas de vigilância de Londres.⁷⁰ O londrino médio é filmado 300 vezes por dia, o que atribui à cidade o título de “capital mundial da cctv” (*closed-circuit television*). A ong de direitos humanos Big Brother Watch lidera uma campanha, retratada no documentário, contra o uso da tecnologia pela polícia inglesa, com ações de conscientização e protesto pelas ruas de Londres. Numa das cenas, o grupo de ativistas intercede em defesa de um cidadão assediado e multado pela polícia pelo simples fato de ter passado em frente a uma das câmeras com o rosto coberto. Em outra cena, um rapaz negro, por reconhecimento equivocado do sistema, é revistado pela polícia na frente dos amigos.

O argumento legítimo da ong é que a lei inglesa, como nos demais países, protege os dados pessoais de identificação, exigindo consentimento ou autorização judicial para a coleta de dna e impressão digital. A *Protection of Freedoms Act 2012: DNA and Fingerprint (Lei de Proteção da Liberdade de 2012: disposições sobre dna e Impressão Digital)*, em vigor desde outubro de 2013, prevê

que apenas os condenados por crime tenham seus registros de impressão digital e perfil de dna retidos indefinidamente, os demais devem ser destruídos no prazo máximo de seis meses após a coleta. Contrariando essa lei e a lei de proteção de dados europeia de 2018 (gdpr), as câmaras de vigilância captam dados biométricos à revelia do cidadão, com zero transparência.

Outro exemplo apresentado no documentário é o uso de reconhecimento facial pelo fbi. A agência treina seus algoritmos em dados coletados de registros estaduais de motoristas habilitados, sem mandato e sem medida de proteção; a coleta já ocorreu em 18 estados norte-americanos. Essas ações são inconstitucionais. Como mostram os filmes e as séries, para entrar na casa de um suspeito, a polícia precisa de consentimento do morador ou de mandato judicial, e as provas colhidas ilegalmente não são aceitas no tribunal.

O documentário clama por regulamentação. Cathy O’Neil, por exemplo, propõe a criação de uma agência reguladora da inteligência artificial, que exigiria dos desenvolvedores da tecnologia evidências de que seus sistemas não causem danos à sociedade. “Prove que é legal antes de lançar”, vaticina O’Neil. A prática, contudo, está mostrando que identificar e denunciar os danos é mais fácil do que regulamentá-los.

O estado de Massachusetts, em resposta ao movimento iniciado em 2019 por Kade Crockford, ativista da American Civil Liberties Union (aclu), aparentemente, conseguiu encontrar um equilíbrio na regulamentação do reconhecimento facial, permitindo que os agentes da lei aproveitem seus benefícios enquanto constroem proteções aos cidadãos. O projeto, que entra em vigor em julho deste ano, prevê que a polícia tem de obter a permissão de um juiz antes de fazer uma pesquisa de reconhecimento facial, e em seguida ter alguém da polícia estadual para realizar a busca; um oficial local não pode simplesmente baixar um aplicativo de reconhecimento facial e fazer uma pesquisa. Em paralelo, foi definida uma comissão para estudar as políticas de reconhecimento facial e fazer recomendações, por exemplo, se um réu criminal deve ser informado de que foi identificado usando-se a tecnologia.

A proposta de auditar os sistemas de inteligência artificial vem sendo debatida em vários fóruns. A convite do governo francês, o matemático Cédric Villani chefiou uma força-tarefa sobre a estratégia de ia para a França e a Europa com a missão de avaliar a consistência dos modelos em relação aos princípios e normas vigentes; ao final, Villani recomendou a obrigatoriedade de auditar os sistemas autônomos. A ideia de auditoria é igualmente defendida pelo filósofo Luciano Floridi; contudo, o próprio Floridi aponta um conjunto de restrições conceituais, técnicas, econômicas, sociais, organizacionais e institucionais para sua implementação.⁷¹

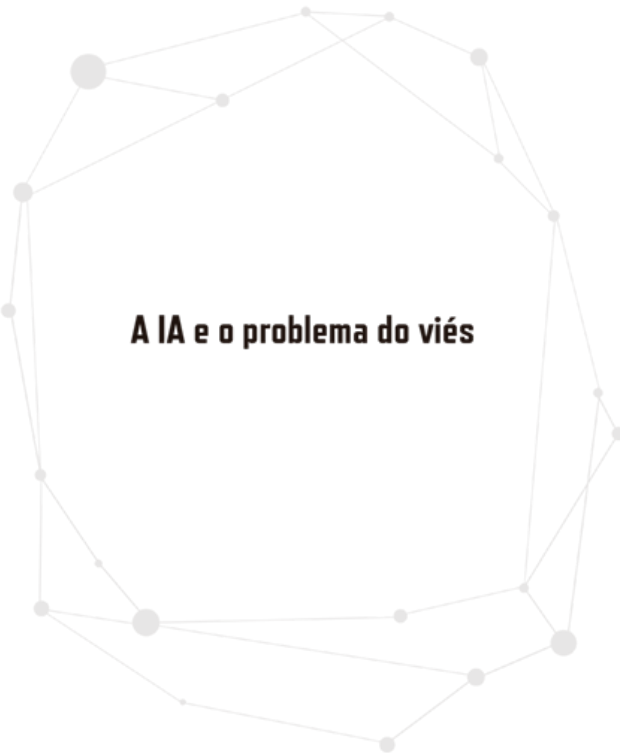
Complicando ainda mais, os sistemas de inteligência artificial são dinâmicos, transformam-se com novos dados; não resolve, portanto, auditar apenas na partida. A velocidade e a descentralização no desenvolvimento da

tecnologia dificultam replicar o arcabouço regulatório da indústria farmacêutica (mais concentrada, relativamente mais fácil de monitorar e fiscalizar). Os algoritmos de ia são, em geral, proprietários, ou seja, são protegidos por sigilo comercial; as tecnologias são sofisticadas, demandam conhecimento especializado que, em geral, escapam aos reguladores e legisladores; e a natureza do órgão de auditoria é crítica, cada opção tem seus problemas (órgão de governo, órgão multilateral, terceirizado-privado).

A Comissão Europeia anunciou que vai propor uma regulamentação para inteligência artificial ainda no primeiro trimestre de 2021, com o objetivo de salvaguardar os valores e os direitos fundamentais da União Europeia e a segurança dos usuários. O projeto é resultado de um longo processo que contou com a contribuição de 52 especialistas e permaneceu em consulta pública durante todo o ano 2020. Os pontos antecipados ainda são bem gerais – garantir supervisão humana e informações claras sobre capacidades e limitações dos sistemas.

A Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco) elaborou um documento sob consulta pública igualmente generalista, mais próximo de uma carta de intenções; acompanhando a tendência, não contempla as limitações da tecnologia nem as restrições mencionadas anteriormente.⁷² Sobre a Portaria gm n.º 4.617, do Ministério da Ciência, Tecnologia e Inovações do Brasil, de 6 de abril de 2021, que institui a “Estratégia Brasileira de Inteligência Artificial e seus eixos temáticos”, ver coluna do jornalista Ronaldo Lemos.⁷³

Denunciar os danos dos sistemas de inteligência artificial é mandatório, como enfrentá-los é complexo, o desafio é migrar de princípios e intenções gerais para diretrizes e regulamentações.



A IA e o problema do viés

Um sistema de IA é considerado enviesado, ou seja, apresenta viés nos resultados, quando ocorre discriminação sistemática contra certos indivíduos ou grupos de indivíduos com base em certos traços ou características, particularmente, atributos sensíveis, como gênero, etnia, raça. No século XVI o termo viés adquire o significado de “preconceito” e, por volta de 1900, no campo da estatística, o significado técnico como diferenças sistemáticas entre uma amostra e a população-universo. No campo do direito, o viés refere-se a uma opinião preconcebida, um julgamento fundado em preconceitos, o oposto de uma decisão proveniente de avaliação imparcial dos fatos.

Em geral, atribui-se o viés integralmente às bases de dados tendenciosas. Porém, o viés pode emergir antes da coleta de dados em função das decisões tomadas pelos desenvolvedores (os atributos e variáveis contemplados no modelo, inclusive, determinam a seleção dos dados). No caso de viés associado aos dados, existem duas principais origens: os dados coletados não representam a composição proporcional do universo objeto em questão, ou os dados refletem os preconceitos existentes na sociedade. O primeiro caso pode ocorrer se uma base de dados de treinamento privilegiar uma categoria específica, seja ela majoritária ou minoritária, por exemplo, algoritmos de reconhecimento facial treinados em base de dados predominantemente de homens de pele clara. O segundo caso pode ser consequência do enviesamento da própria realidade, o que seria um viés histórico, por exemplo, algoritmos de seleção de candidatos para funções de tecnologia treinados em base de dados na qual predominam currículos de profissionais do sexo masculino.

Resultados tendenciosos podem decorrer, igualmente, de erros na rotulagem da base de dados, processo que antecede o aprendizado supervisionado, e/ou na própria geração de dados. Por exemplo, a não desagregação dos dados por gênero, a maior incidência no banco de dados imagético de imagens originadas por brancos (maior poder aquisitivo, mais acesso a dispositivos sofisticados, maior a probabilidade de gerar imagens de melhor qualidade), a desproporção de bases de dados originadas nos Estados Unidos comparativamente a países orientais, como China e Índia, que representam parte relevante da população mundial.

O viés surge, igualmente, da falta de diversidade nas equipes de desenvolvedores de ia, a maioria composta de cientistas e engenheiros da computação, homens brancos de países ocidentais, na faixa etária entre 20 e 40 anos, com experiências, opiniões e preconceitos similares. Outro aspecto a ser considerado é a variação de desempenho dos sistemas entre o ambiente em que foi treinado e testado e o ambiente do mundo real.

Especialistas em IA estão empenhados em identificar formas de eliminar ou, ao menos, mitigar os vieses dos modelos a partir de variadas abordagens, mas ainda sem muito êxito. Em paralelo, emergem iniciativas de autorregulamentação

e de regulamentação da ia.

Este bloco tem cinco artigos, o primeiro e o terceiro tratam do viés nos sistemas de reconhecimento facial, o segundo, do conceito e da prática da ética *by design*, o quarto é sobre a lacuna de dados de gênero, e o quinto é sobre o viés na interação homem-máquina por voz.

Sistemas de vigilância com reconhecimento facial: escrutínio e reação das gigantes de tecnologia

19.6.2020

Em artigo publicado em abril último pela editora Gallimard, como parte da série *Tracts de crise (Fôlders de Crise, 21-04-2020)*, intitulado em português “Um festival de incerteza”, o pensador francês Edgar Morin nos oferece uma comovente e envolvente reflexão sobre a crise da covid-19.⁷⁴ Entre os temas abordados está a ciência, que, segundo ele, “não tem um repertório de verdades absolutas, e suas teorias são biodegradáveis sob efeito de novas descobertas”. Morin propõe um debate sobre o antagonismo entre prudência e urgência, e seus riscos intrínsecos, alertando para os processos contraditórios que surgem em situações de crise: um estimula a imaginação e a criatividade em busca de novas soluções, e outro se apegava ao passado, tentando minar o novo.

Pela sua natureza disruptiva, e também por estarem em seus primórdios, com limitações de várias naturezas, as tecnologias de inteligência artificial estão sob escrutínio intenso, entre elas os sistemas de vigilância com reconhecimento facial.

As técnicas de processamento de imagens (redes neurais/*deep learning*) são aplicadas desde a identificação de imagens captadas por drones e satélites da superfície da Terra até imagens de tomografia para diagnóstico de contaminação por covid-19, passando pela detecção de pragas na agricultura, pelo monitoramento e personalização no trato com o gado, além do uso em plataformas *online*, redes sociais e aplicativos. O mais controverso, contudo, tem sido o uso nos sistemas de vigilância. Segundo o Ada Lovelace Institute – órgão independente cuja missão é garantir que a ia funcione a favor da sociedade –, testes da tecnologia de reconhecimento facial em contextos de policiamento no Reino Unido relataram mais de 90% de ocorrências incorretas.

Em dezembro de 2019, o relatório *Face Recognition Vendor Test (frvt)*, publicado pelo Instituto Nacional de Padrões e Tecnologia dos Estados Unidos (National Institute of Standards and Technology, nist), descreve e quantifica os diferenciais demográficos para 200 algoritmos de reconhecimento facial de quase 100 desenvolvedores, envolvendo mais de 18 milhões de imagens de mais de oito milhões de pessoas.⁷⁵ O relatório encontrou evidências empíricas da existência de diferenças demográficas na maioria dos algoritmos avaliados, com os resultados afetados por viés de etnia, idade e sexo: afro-americanos e nativos americanos ilhéus do Pacífico foram identificados erroneamente, bem como crianças e idosos.

Em alguns casos, os asiáticos e afro-americanos foram identificados com erro até 100 vezes mais do que os homens brancos. As maiores taxas de precisão

foram encontradas entre homens brancos de meia-idade. O nist alertou as autoridades federais sobre os riscos do uso como aplicação da lei de segurança nacional em áreas críticas, pedindo expressamente que o governo reconsiderasse os planos de utilizar essas tecnologias na proteção de fronteiras.

O relatório do nist é mais um elemento de pressão contra as falhas desses sistemas, pressão que começa a afetar as gigantes de tecnologia. Em junho último, ibm, Microsoft e Amazon anunciaram a limitação do uso de seus sistemas de reconhecimento facial pelos departamentos de polícia (nenhuma das empresas se pronunciou sobre o uso pelo serviço de Imigração e Alfândega).

Em 8 de junho, a ibm declarou, em carta ao Congresso dos Estados Unidos, que não oferecerá mais *software* de análise ou reconhecimento facial de uso geral, e não mais desenvolverá ou pesquisará a tecnologia. “A ibm se opõe firmemente e não tolerará o uso de nenhuma tecnologia de reconhecimento facial, incluindo tecnologia de reconhecimento facial oferecida por outros fornecedores, para vigilância em massa, perfil racial, violações dos direitos e liberdades humanos básicos, ou qualquer finalidade que não seja consistente com nossos valores e princípios de confiança e transparência. Acreditamos que agora é a hora de iniciar um diálogo nacional sobre se, e como, a tecnologia de reconhecimento facial deve ser empregada pelas agências policiais nacionais”, diz trecho da carta assinada por Arvind Krishna, ceo da ibm.

Em 10 de junho, a Microsoft anunciou que deixará de vender a tecnologia aos departamentos de polícia até que seja promulgada uma lei federal com regras de uso, ressaltando seu apoio anterior à legislação californiana sobre a proteção à privacidade de dados pessoais que inclui restrições ao uso policial.

Também em 10 de junho, a Amazon, maior fornecedora da tecnologia para fins de aplicação de leis – sua subsidiária Ring, por exemplo, estabeleceu parcerias com mais de 1.300 agências policiais para usar imagens de suas câmeras de segurança doméstica em investigações criminais –, surpreendeu ativistas e pesquisadores de direitos civis quando anunciou que colocaria uma moratória de um ano no uso da Rekognition pela polícia. A decisão resulta de dois anos de pesquisa e pressão externa demonstrando as falhas técnicas da Rekognition, não apenas por instituições externas, como a Fundação aclu de Washington, que recolheu mais de 150 mil assinaturas de petições contra o sistema, mas também pelos próprios funcionários da Amazon, que ecoaram num memorando interno as preocupações e protestos externos.

Alguns fornecedores, cientes de suas limitações, têm se empenhado em aperfeiçoar os sistemas, particularmente a questão do viés, como é o caso da ibm. Em janeiro de 2019, a empresa lançou um conjunto de dados chamado Diversity in Faces, com mais de um milhão de imagens de rostos, mas a iniciativa fracassou, ao ser descoberto que as imagens haviam sido retiradas da plataforma Flickr sem consentimento dos usuários, provocando novas discussões sobre como treinar eticamente os sistemas de reconhecimento facial.

Retomando Edgar Morin, a urgência em derrotar a covid-19 comprometeu a prudência de parte dos pesquisadores em estudos e modelos publicados em revistas científicas, algumas com revisões por pares. No caso dos sistemas de vigilância com tecnologias de reconhecimento facial, contudo, não se justifica comprometer a prudência porque não existe urgência.

O Facebook anunciou, em 15 de junho, o lançamento, no Brasil, do Facebook Pay, seu novo sistema de pagamentos pelo WhatsApp. Alguns dias depois, dia 23, o Banco Central e o Conselho Administrativo de Defesa Econômica (Cade) determinaram sua suspensão temporária. Entre as preocupações das autoridades brasileiras estão a proteção dos dados pessoais gerados no aplicativo, incluindo seu potencial compartilhamento com as três redes sociais controladas pela corporação (Facebook, WhatsApp e Instagram), e a garantia de que não haja vantagem indevida, que comprometa a livre concorrência, frente ao número expressivo de seus usuários (consequentemente, de dados). Segundo os envolvidos, o projeto levou dois anos para ficar pronto, mas, pelas informações divulgadas, aparentemente, não atentou na íntegra para as questões éticas associadas ao uso de dados pessoais.

O modelo econômico que tende a prevalecer no século xxi, orientado por dados (*data-driven economy*), essencialmente, extrai *inputs* de grandes conjuntos de dados com técnicas estatísticas de inteligência artificial, com o objetivo de personalizar, prever, automatizar e otimizar processos, produtos e serviços. Sendo os dados pessoais a base desses modelos de negócio, torna-se mandatório contemplar as leis de proteção de dados – lgpd brasileira, gdpr europeia e ccpa californiana – que, implicitamente, recomendam a adoção do conceito de *privacy by design*.

Em geral, os cientistas de tecnologia, seja na academia ou no mercado, não atentam para os impactos éticos e sociais dos sistemas que desenvolvem, tornando mais complexa a identificação, a regulação e a fiscalização *a posteriori*. Essa suposta ausência de sensibilidade para as questões éticas esteve no cerne de uma polêmica que agitou a comunidade de inteligência artificial, ao envolver alguns de seus renomados representantes, entre eles o chefe de ia do Facebook, Yann LeCun, vencedor do Prêmio Turing como um dos pioneiros da técnica de redes neurais profundas (*deep learning*).

A polêmica explodiu na conferência *Computer Vision and Pattern Recognition* (cvpr), realizada virtualmente em junho último. Polemizando em primeiro plano com LeCun esteve a cientista da computação etíope-americana Timnit Gebru, colíder técnica da equipe de inteligência artificial ética do Google, cujas pesquisas sobre o preconceito de raça e gênero nos sistemas de reconhecimento facial têm subsidiado os congressistas norte-americanos a favor da proibição pelo governo federal desses sistemas, e influenciou, igualmente, as decisões da ibm, da Microsoft e da Amazon de interromper seu uso pela polícia.⁷⁶

A polêmica começou na semana anterior à conferência, quando um observador introduziu no pulse uma foto do ex-presidente Obama, e o sistema gerou uma foto de um homem branco; na sequência, o sistema transformou Muhammad Ali em branco, traços de mulheres asiáticas em traços de mulheres brancas, dentre outras distorções claramente de viés racial. O pulse é um modelo de visão computacional desenvolvido por pesquisadores da Universidade de Duke, utilizando redes gans (*generative adversarial networks*), um tipo de arquitetura de aprendizado de máquina, e cujo algoritmo (Stylegan/Nvidia) foi treinado com dados do Flickr Face hq; ele cria rostos humanos assustadoramente realistas usados em perfis falsos de mídia social.

O estopim da polêmica foi a declaração de Yann LeCun a essas evidências: “Os sistemas de ml (*machine learning*) são tendenciosos quando os dados são tendenciosos”,⁷⁷ imediatamente contestada por diversos especialistas, que o acusaram de minimizar o problema ou até mesmo de se isentar da responsabilidade (lembrando que o tema do racismo está na pauta nos Estados Unidos neste momento, com a ascensão do movimento Black Lives Matter). Timnit Gebru convidou LeCun para assistir a seu tutorial gpr – apresentado na abertura da Conferência cvpr –, cuja mensagem central é que o viés nos modelos de inteligência artificial não pode ser imputado exclusivamente aos dados. A polêmica se estendeu por dias nas redes sociais, principalmente no Twitter, alcançando tons não recomendados entre pesquisadores.

Diante da repercussão negativa, o chefe de inteligência artificial do Google, Jeff Dean, pediu que a comunidade reconhecesse que o viés vai além dos dados, e o vice-presidente de ia do Facebook, Jerome Pesenti, desculpou-se, afirmando ser importante ouvir as experiências de pessoas que sofreram injustiça racial.⁷⁸

Do ponto de vista estritamente técnico, a resposta de LeCun é em parte correta, percentual significativo do preconceito e da injustiça nos sistemas de inteligência artificial vem da base de dados usada no desenvolvimento do modelo, que podem ser, inclusive, amplificados no processo posterior de treinamento dos algoritmos. As recomendações desses sistemas inteligentes são efetivamente influenciadas pelos dados tendenciosos, por isso recomenda-se que a base de dados utilizada seja criteriosamente avaliada.

A forte reação de alguns membros importantes da comunidade de ia deve-se mais à forma sucinta da resposta; sendo ele um dos líderes da área, esperava-se que fizesse ponderações mais amplas, alertando, por exemplo, sobre as limitações e distorções dos sistemas, e reconhecesse que, efetivamente, falta desenvolvimento técnico para mitigar o potencial viés. Esperava-se, igualmente, uma recomendação aos desenvolvedores sobre a necessidade de um exame mais cuidadoso da base de dados em busca de sinais de viés ou, como alternativa, sobre treinar os algoritmos usando conjuntos maiores e mais diversificados de dados. Por último, esperava-se que LeCun aconselhasse os desenvolvedores a explicitarem que se trata de modelos estatísticos de probabilidade, logo seus

resultados nunca são 100% assertivos (independentemente da base de dados). Ou seja, que ele se portasse como um líder da comunidade e fizesse jus a função que desempenha em uma das *big techs*.

O conceito de “ética *by design*” deve ser adotado desde o início do desenvolvimento dos sistemas de IA, não só pelas implicações éticas do uso de dados pessoais, como também pelos impactos sociais em áreas sensíveis à sociedade, como trabalho, educação e saúde. Daí a importância de formar equipes multidisciplinares com pesquisadores de ciências exatas e humanas, oferecendo à sociedade soluções mais amigáveis, menos ameaçadoras, que construam igualdades e não desigualdades.

Surpreendentemente, Londres tem mais câmeras de vigilância do que Pequim

19.2.2020

Londres é a cidade mais bem avaliada no iese Cities in Motion (2020), índice de avaliação de cidades da Business School of the University of Navarra (IESE), Espanha. Prestigiado internacionalmente, o índice é composto de 96 indicadores, cobrindo 174 cidades de 80 países (79 capitais). Londres, contudo, é também a terceira cidade mais vigiada do mundo; entre as primeiras 10, nove são chinesas. Pequim ocupa a quinta posição, ou seja, a capital da China, país do sistema de crédito social (*Social Credit System*), tem menos câmeras de vigilância proporcionalmente à população do que a cidade europeia: são 1,15 milhões de câmeras para uma população de 20 milhões de pessoas em Pequim (56,20 câmeras a cada mil habitantes), contra 627.727 câmeras para uma população de 9,3 milhões de pessoas em Londres (67,47 câmeras a cada mil habitantes).⁷⁹

O londrino médio é filmado 300 vezes por dia, o que atribui à cidade o título de “capital mundial da cctv” (*closed-circuit television*). O Estado controla uma parte menor do sistema de vigilância: estudo da British Security Association indicou uma proporção entre câmeras privadas e públicas de 70 para 1. Os números tendem a ser maiores porque as câmeras domésticas não precisam de registro ico (Information Commissioner’s Office), só as câmeras de empresas. Em média, teoricamente as gravações são armazenadas por duas semanas.

Ampliando ainda mais a vigilância, apesar das falhas identificadas em anos de testes, em janeiro de 2020, o Departamento de Polícia de Londres anunciou a incorporação de tecnologias de reconhecimento facial em suas câmeras. Embora o objetivo seja localizar suspeitos de terem cometido crimes graves, a medida provocou forte reação de grupos contrários à expansão do “Estado de Vigilância”, como o Big Brother Watch; o foco dos protestos é a ausência de debate público e de regulamentação apropriada. Moradores de cidades com histórico de ataques terroristas como Londres, contudo, tendem a ser mais benevolentes ao dilema proteção e segurança *versus* liberdades civis.

Como amplamente denunciado no âmbito das manifestações do Black Lives Matter, em maio de 2020, as tecnologias de reconhecimento facial têm viés de etnia, idade e sexo.⁸⁰ Diversos escrutínios desses sistemas, por parte de centros de pesquisa, institutos e grupos de direitos civis, encontraram evidências de falhas graves, como a que ocorreu no Rio de Janeiro: em janeiro de 2018, a polícia carioca contratou o sistema britânico Facewatch com o propósito de identificar 1.100 criminosos foragidos, que seriam reconhecidos ao cruzar uma das câmeras de vigilância espalhadas pela cidade; em 2019, uma mulher foi detida em

Copacabana e encaminhada à delegacia, onde foi confirmado que não se tratava da criminosa procurada (que, inclusive, estava presa na ocasião).

O reconhecimento facial é uma das aplicações da técnica de aprendizado de máquina, subcampo da inteligência artificial, denominada *deep learning* (aprendizado profundo). Considerada uma das áreas mais bem-sucedidas da ia, serve a múltiplas tarefas: desde o simples reconhecimento de imagens em pesquisas no Google até a interpretação de tomografias e a biometria facial, substituindo as tradicionais senhas (nem sempre seguras nem fáceis de memorizar).

O fundamento dessa técnica é que o algoritmo “aprende” com base em exemplos extraídos dos dados. Para que seus modelos atinjam desempenhos aceitáveis, precisam ser treinados em extensas bases de dados, ou seja, a precisão do resultado é proporcional ao número e à qualidade dos exemplos contemplados no modelo (exemplos contidos nos dados). Seus algoritmos são capazes de lidar com dimensões da ordem de grandeza de milhões, por exemplo, milhões de *pixels* numa imagem. Embora denominadas “redes neurais”, pela inspiração no funcionamento do cérebro, o sistema visual humano transcende em muito o simples reconhecimento do objeto: nosso sistema visual é capaz de compreender cenas inteiras, ou seja, captar múltiplos objetos de uma cena e a relação entre esses objetos, e processar informações em 3D.

A resistência ao uso dos sistemas de reconhecimento facial para vigilância decorre, fundamentalmente, de duas limitações intrínsecas à técnica: o viés e a opacidade (não explicabilidade). O viés decorre, entre outras origens, da base de dados utilizada no treinamento do modelo, que pode ser tendenciosa, ao não refletir a diversidade do universo considerado. Se a base de dados for predominantemente de homens brancos, o sistema terá dificuldade de reconhecer com precisão, por exemplo, mulheres negras. Um exemplo mais simples: um veículo autônomo treinado com dados de uma cidade norte-americana não apresentaria o mesmo desempenho nas ruas de uma cidade inglesa, onde a direção é no lado esquerdo. A base de dados pode ser enviesada por refletir os preconceitos dos humanos contidos nos dados.

A opacidade decorre do desconhecimento de como os chamados dados de entrada (*inputs*) geram o dado de saída (*output*), ou seja, como o sistema correlaciona os exemplos contidos nos dados. Com enormes recursos computacionais e muitos *terabytes* de dados, o número de recursos possível de incluir em um sistema de reconhecimento facial ultrapassa o nível de compreensão de um humano racional (incompatibilidade entre o processo matemático do aprendizado de máquina com a cognição em escala humana).

Existem meios de mitigar as limitações da técnica, o que, por um lado, requer um empenho dos desenvolvedores desses sistemas e, por outro, uma maior familiaridade dos usuários para criticar, checar e ponderar sobre os resultados.

Lacuna de dados de gênero: mulheres invisíveis num mundo projetado por homens

2.4.2021

O título da coluna reproduz o título do livro *Invisible Women: Exposing Data Bias in a World Designed for Men* (*Mulheres invisíveis: expondo o viés dos dados num mundo projetado por homens*), da escritora, jornalista e ativista feminista britânica Caroline Criado Perez. Considerado o Business Book of The Year 2019 pelo *Financial Times* e pela McKinsey, o livro agrega valiosa contribuição para o debate global sobre discriminação de gênero nos modelos baseados em inteligência artificial. Segundo a autora, a sub-representação de 50% da população nos bancos de dados implica um registro enviesado da história humana.⁸¹

Criado Perez realiza um extenso levantamento histórico da invisibilidade feminina. Para a autora, a tendência universal de considerar o homem como “padrão humano” gera um viés de gênero nos dados, preserva automaticamente a desigualdade e compromete o critério de objetividade. “A partir da teoria do homem caçador, os cronistas do passado deixaram pouco espaço para o papel das mulheres na evolução da humanidade, seja cultural ou biológica”, pondera a autora.

Como a técnica de inteligência artificial que permeia a maior parte das aplicações atuais é baseada em dados (*machine learning/deep learning*), a sociedade está tomando decisões enviesadas por gênero em número muito maior do que o percebido. Na Inglaterra, por exemplo, as mulheres têm 50% mais chances de ser diagnosticadas erroneamente após um ataque cardíaco, em função da predominância dos homens nos estudos científicos sobre insuficiência cardíaca.

No combate à covid-19, a não coleta de dados desagregados por sexo, ao não contemplar as distinções sexuais na função imunológica, impacta negativamente a identificação de sintomas, taxas de contaminação e de mortalidade. A taxa de mortalidade da covid-19 é de 2 para 1 entre homens e mulheres; sem coleta desagregada, não é possível identificar a razão, nem ao menos saber se os homens têm mais probabilidade de contrair covid-19 ou mais probabilidade de morrer de covid-19.

Criado Perez alerta que, no final de março de 2020, apenas seis dos 20 países mais afetados pela covid-19 estavam publicando dados desagregados por sexo, e os Estados Unidos e o Reino Unido só o fizeram plenamente em maio. Em setembro de 2020, apenas 30% dos países relataram dados desagregados por sexo em relação a contaminação e morte, e menos de 50% dos países desenvolvidos publicaram dados desagregados. Ilustrando a importância da

desagregação, um estudo realizado em 2016, num hospital em Long Island, em Nova York, correlacionou o hormônio feminino estrogênio com resultados positivos no combate aos vírus em geral; em 2020, na tentativa de salvar vidas, esse mesmo hospital chegou a injetar estrogênio em seus pacientes homens com covid-19 (os resultados não foram apurados plenamente, ou não são públicos).

A lacuna de dados de gênero está presente também nos estudos climáticos. Segundo Criado Perez, até 2007, ano de publicação da primeira pesquisa com desagregação por gênero, não existiam dados sobre a disparidade entre homens e mulheres na mortalidade por desastres naturais: dados de 141 países, entre 1981 e 2002, revelaram que as mulheres têm mais probabilidade de morrer em desastres naturais do que os homens. As causas são culturais e comportamentais: os homens indianos, por exemplo, têm maior probabilidade de sobreviver a terremotos noturnos, porque dormem do lado de fora e nos telhados nas noites quentes, o que é interdito às mulheres. No Sri Lanka, outro exemplo, aprender a nadar e a escalar são prerrogativas dos homens: o tsunami de dezembro de 2004 matou quatro vezes mais mulheres do que homens.

O viés de gênero reflete, em parte, a não diversidade das equipes desenvolvedoras de tecnologia: as mulheres representam apenas 11% dos desenvolvedores de *software*, 25% dos funcionários do Vale do Silício e 7% dos sócios em empresas de capital de risco. A diversidade não é apenas uma questão ética e moral, mas tem vários efeitos inclusive sobre a ciência: análise de 1,5 milhão de artigos científicos publicados entre 2008 e 2015 constatou que a probabilidade de um estudo envolver análise de gênero e sexo correlaciona-se com a proporção de mulheres entre seus autores, efeito maior se uma mulher for líder do grupo de autores.

A Unesco, em 2019, em parceria com o governo da Alemanha e a EQUALS Skills Coalition (ramo da EQUALS Global Partnership, dedicada a capacitar mulheres e meninas na aquisição de habilidades em tecnologia), publicou o estudo *I'd Blush, If I Could*, que aborda a lacuna de gênero nas habilidades digitais e compartilha estratégias para reduzir essa lacuna por meio da educação. O título do estudo reproduz a resposta-padrão do assistente virtual Siri a um insulto: “Eu ficaria corada, se pudesse”.⁸² O estudo observou, paradoxalmente, que os países com os níveis mais altos de igualdade de gênero, como os países europeus, têm taxas mais baixas de mulheres na pós-graduação em Ciências da Computação e campos relacionados; e os países com baixos níveis de igualdade de gênero, como os países árabes, têm as maiores proporções de mulheres em cursos de tecnologias avançadas.

Predominam nos assistentes virtuais nomes e vozes femininos, como a Alexa, da Amazon, a Siri, da Apple, a Cortana, da Microsoft, e, ainda pior, a postura desses assistentes é submissa: o relatório da Unesco constatou, por exemplo, que quando um usuário diz à Alexa “Você é gostosa”, a resposta automática é: “É legal da sua parte dizer isso!”. A codificação dos preconceitos em

produtos de tecnologia perpetua o preconceito de gênero da sociedade. “Como a fala da maioria dos assistentes de voz é feminina, isso envia um sinal de que as mulheres são ajudantes prestativas, dóceis e ansiosas por agradar, disponíveis ao toque de um botão ou a um comando de voz direto como ‘Hey’ ou ‘OK’”, pondera o relatório.

Os assistentes virtuais não têm poder de ação, honram comandos e respondem a perguntas independentemente de seu tom ou hostilidade, reforçando os preconceitos de gênero comumente aceitos de que as mulheres são subservientes e tolerantes a um tratamento inadequado. A Unesco adverte que a presença desses assistentes virtuais nos lares mundo afora tem o potencial de influenciar as interações com mulheres reais, e ressalta: “Quanto mais essa cultura ensinar as pessoas a igualar as mulheres às assistentes, mais as mulheres reais serão vistas como assistentes – e penalizadas por não serem como as assistentes”.

O relatório repercutiu intensamente na mídia, com artigos publicados em jornais de grande circulação, como *The New York Times*,⁸³ *The Guardian*,⁸⁴ *Le Monde*,⁸⁵ *El País*,⁸⁶ *O Globo*,⁸⁷ entre outros. Provavelmente como reação ou mera coincidência, a Apple anunciou que a partir do iOS 14.5 o usuário poderá escolher a voz da Siri ao se cadastrar no sistema.

Allison Gardner, cofundadora da Women Leading in ai, organização dedicada a promover a diversidade e a boa governança em IA, reconhece que nem sempre o preconceito é malicioso – em geral resulta da falta de consciência de que o preconceito existe – e atribui parte da causa à não diversidade das equipes de desenvolvedores, uma das barreiras da ética *by design*.

A interação humano-máquina por voz é potencialmente inclusiva, mas na prática tem sido discriminatória

3.9.2021

O jornal *The Washington Post*, em 2018, reuniu um grupo de pesquisadores para investigar o efeito de distintos sotaques na interação com sistemas de voz baseados em inteligência artificial; foram testados milhares de comandos de voz ditados por mais de 100 pessoas em quase 20 cidades norte-americanas. Os resultados indicaram disparidades relevantes: quando Meghan Cruz diz “*Hi, Alexa*”, sua Alexa Amazon oferece de imediato a resposta solicitada, contudo, quando Andrea Moncada, estudante universitária criada na Colômbia, diz o mesmo em seu leve sotaque espanhol, a Alexa nem se manifesta.

A ativação por voz tende a prevalecer na interface com os dispositivos digitais, “mas para pessoas com sotaque os sistemas inteligentes de voz podem parecer desatentos e indiferentes. Para muitos norte-americanos, a onda do futuro tem um problema de discriminação e os está deixando para trás”, conclui o estudo.⁸⁸ Ou seja, uma tecnologia potencialmente inclusiva, em particular no Brasil, onde cerca de 30% da população é classificada como analfabeta funcional, está sendo excludente na prática.

Estudos realizados por pesquisadores da Universidade de Stanford, com base em dispositivos da Amazon, ibm, Google, Microsoft e Apple, acusaram que os sistemas de reconhecimento de voz automatizados com inteligência artificial (*automatic speech recognition*, asr), em média, apresentam taxa de erro de palavra de 0,35 para falantes negros em comparação com 0,19 para falantes brancos.⁸⁹ Os efeitos são perversos, considerando-se a variedade de aplicativos de conversão da linguagem falada em texto: assistentes virtuais, preenchimento de registros médicos por voz, tradução automática, entrevistas *online* automatizadas nos processos de seleção de candidatos a emprego, entre inúmeros outros aplicativos.

Nesse estudo, os sistemas foram testados com mais de 2 mil amostras de entrevistas gravadas com afro-americanos e brancos – amostras de fala negra foram retiradas do Corpus of Regional African American Language (CORAAL), e as de fala branca, de entrevistas conduzidas pelo Voices of California. O estudo constatou que essas disparidades étnicas são resultantes das características fonológicas, fonéticas ou prosódicas (aspectos dos sons da fala, como acentuação e entonação) do inglês afro-americano, e não das características gramaticais ou lexicais. Uma limitação do estudo é que as amostras de áudio de falantes brancos e negros vieram de diferentes áreas geográficas do país – as primeiras, coletadas na Califórnia, e as últimas, no leste dos Estados Unidos –, logo é possível que algumas das diferenças identificadas sejam consequência da variação linguística

regional, e não étnica.

As variações de sotaque, dizem os engenheiros, representam um dos maiores desafios para as empresas que trabalham no desenvolvimento de sistemas de reconhecimento de voz. Um grupo de pesquisadores brasileiros – Lanna Lima, Elizabeth Furtado e Vasco Furtado, da Universidade de Fortaleza, e Virgílio Almeida, da Universidade Federal de Minas Gerais – analisou a presença de vieses na interação via áudio, por meio de um experimento da Universidade de Fortaleza, conduzido em laboratório com 20 voluntários interagindo com o Google Assistant e a Siri, da Apple.⁹⁰ “O estudo apresenta uma análise preliminar, indicando que o processo de formação de assistentes de *smartphones*, para o português brasileiro, pode ser tendencioso para vozes de indivíduos da parte mais desenvolvida do país”, concluem os autores.

A origem da disparidade está principalmente na base de dados tendenciosa utilizada para treinar esses sistemas (predominantemente, norte-americanos brancos), quando deveria refletir a diversidade de sotaques e dialetos, ou seja, representar a composição proporcional do universo objeto em questão (diversificado em relação às subpopulações). Considera-se que existe um enviesamento na base de dados quando o sistema exibe um erro sistemático no resultado (“enviesamento estatístico”). O desempenho dos sistemas varia quando eles saem do ambiente controlado em que são treinados e testados, com dados “higienizados” dos laboratórios, e passam a usar os dados do mundo real; essas diferenças impactam fortemente a acurácia dos resultados.

O viés, contudo, também pode emergir de decisões tomadas pelos desenvolvedores (os atributos e variáveis contemplados no modelo, inclusive, determinam a seleção dos dados). No desenvolvimento de um sistema de IA com a técnica de redes neurais profundas (*deep learning*), a tarefa inicial dos cientistas da computação é identificar o problema a ser resolvido, em qual situação e com qual objetivo o sistema será utilizado; o passo seguinte é traduzir esse problema a ser resolvido em variáveis que possam ser computadas (hiperparâmetros). Identificar a influência da subjetividade humana na elaboração do sistema e na configuração do algoritmo de inteligência artificial não é trivial, além de não ser possível eliminá-la completamente, mesmo se e quando identificada.

O Alan Turing Institute sinaliza como um problema crítico a postura dos desenvolvedores e designers de algoritmos de ia que permite que os vieses sistêmicos se infiltrem nos dados. Na perspectiva do instituto, o problema deriva da não priorização de ações para identificar e corrigir desequilíbrios potencialmente discriminatórios por parte dos produtores de tecnologia, em geral, predominantemente homens brancos, logo, isentos dos efeitos adversos de resultados discriminatórios.⁹¹

Brian Christian, no livro *The Alignment Problem: Machine Learning and Human Values*⁹² (*Problema de alinhamento: aprendizado de máquina e valores humanos*), pondera que a última década presenciou “o progresso mais

estimulante, abrupto e preocupante na história do aprendizado de máquina – e, de fato, na história da inteligência artificial. Há um consenso de que uma espécie de tabu foi quebrado: não é mais proibido aos pesquisadores de ia discutir questões de segurança”. Christian identifica duas comunidades distintas preocupadas com as externalidades negativas dos sistemas de inteligência artificial: a primeira, com foco nos riscos éticos atuais da tecnologia, e a segunda, preocupada com os riscos futuros. Ambos os cenários não podem ser abordados dentro dos campos disciplinares tradicionais, mas apenas através do diálogo entre cientistas da computação, cientistas sociais, advogados, reguladores, especialistas em políticas, especialistas em ética. Os efeitos da “discriminação algorítmica” são perversos, contribuem para ampliar a já imensa desigualdade global.

**A economia de dados
e o poder das *big techs***

A informação é um componente essencial no funcionamento dos mercados. Informação de qualidade tende a aumentar a eficiência e reduzir os custos, consequentemente, dota as empresas ricas em dados de vantagem competitiva em relação aos seus concorrentes, aumenta as barreiras à entrada de novos concorrentes e amplia a assimetria informacional entre empresas e entre empresas e usuários/consumidores. A vantagem competitiva está, particularmente, na personalização da comunicação e dos preços de produtos e serviços com base no perfil do usuário/consumidor. A coleta e análise dos dados pessoais permite às empresas identificar quais fatores impulsionam o desejo de compra do consumidor e, inclusive, qual o valor máximo que ele estaria disposto a pagar, processo denominado “discriminação comportamental” (reduz o “excedente do consumidor”, ou seja, o potencial ganho obtido pelo consumidor na comparação de preços).

A hiperconectividade da sociedade, consequentemente, a digitalização da vida (“datificação”), altera os pilares estruturais da economia, atribuindo papel crítico aos dados. Os modelos de negócio baseados em dados (*data-driven business models*) são caracterizados pela capacidade de coletar cada vez mais dados sobre nosso comportamento e nossas preferências, criar perfis sobre nós, personalizar induções e recalibrar com base em nossas interações/respostas. Esses modelos têm o potencial de aumentar o consumo e otimizar a extração de riqueza, contudo, afetam valores valiosos, como privacidade, igualdade e justiça.

Na Economia de Dados, a eficiência corporativa é função do grau de sociabilidade. Na nova lógica, quanto mais interação entre os indivíduos, ou seja, sua sociabilidade e comunicação, mais geração de dados pessoais, mais eficiência operacional.

Neste bloco temos oito artigos que versam sobre as grandes plataformas de tecnologia, os desafios do capitalismo de dados, o dilema entre conveniência e “pesadelo orwelliano”, o significado de Lina Khan como presidente da FTC. O terceiro artigo trata do polêmico documentário *O dilema das redes*.

Tecnologia gratuita em troca de seus dados: o melhor e o pior do capitalismo em um simples *swap*

29.11.2019

A chamada “Economia de Dados”, que tende a ser o modelo econômico predominante no século xxi, caracteriza-se por modelos de negócio baseados em dados (*data-driven business models*), ou seja, na capacidade de identificar nos dados padrões, preferências e hábitos dos usuários/consumidores/clientes. Em geral, as informações são extraídas por meio de modelos de inteligência artificial. O uso de dados pessoais para finalidades comerciais, contudo, é anterior às tecnologias de ia.

O supermercado inglês Tesco, em 1993, lançou o cartão Clubcard. A troca era simples: a cada compra, o cliente registrava o cartão no caixa, recebendo pontos de desconto para uso em futuras compras, e em contrapartida autorizava a coleta de seus dados pessoais (basicamente nome, valor e data da compra). Essa limitada base de dados trouxe *insights* sobre o comportamento de compra dos clientes, promovendo estratégias de marketing inovadoras que contribuíram para transformar a Tesco em líder do setor no Reino Unido.

Na década seguinte, em 2002, a loja de descontos norte-americana Target desenvolveu um modelo estatístico para investigar padrões na base de dados de seus clientes, o que gerou um evento polêmico, frequentemente citado como ilustrativo da ameaça à privacidade do uso dos dados pessoais: a varejista criou um algoritmo para classificar a probabilidade de uma cliente estar grávida; se essa probabilidade ultrapassasse certo limite, em torno de 87%, o sistema enviava automaticamente à cliente cupons de produtos afins (fraldas, hidratantes e loções mais suaves, chupetas). Certo dia, o pai de uma adolescente invadiu uma das lojas da Target, em Minneapolis, reclamando de que sua filha, de apenas 16 anos, tinha recebido os “cupons de gravidez”. O gerente se desculpou de imediato. Tempos depois foi a vez de o pai da adolescente se desculpar: a filha estava efetivamente grávida. Ou seja, a varejista Target soube da gravidez antes de sua família (especula-se que mesmo antes da própria adolescente!). O ocorrido foi narrado em artigo de Charles Duhigg no jornal *The New York Times* e no livro *O poder do hábito*, do mesmo autor.⁹³

Os resultados bem-sucedidos dos modelos de redes neurais – automaticamente “varrem” a base de dados e identificam, através de modelos estatísticos e algoritmos de inteligência artificial, tendências e cenários futuros e a probabilidade de cada um deles acontecer – trouxeram um grau de assertividade inédito aos processos (em alguns casos perto de 90-95%). Mais do que as tecnologias, os dados são os ativos valiosos e, em parte, explicam o poder e o

domínio de mercado das *big techs* (gigantes de tecnologia: Amazon, Apple, Facebook, Microsoft, IBM, Google, Alibaba, Baidu e Tencent).

Um dos efeitos positivos das recentes leis de proteção de dados – a *General Data Protection Regulation* (gdpr), em vigência na Europa desde maio de 2018, e a Lei Geral de Proteção de Dados (lgpd), do Brasil, prevista para entrar em vigor em agosto de 2020 – foi ampliar o debate, consequentemente a visibilidade, sobre o uso dos dados pessoais. Estamos todos mais atentos e resistentes a pedidos não justificados de informações, como cpf em farmácias, telefone e cpf em cadastro na portaria de prédios, entre inúmeras outras solicitações não relacionadas ao “legítimo interesse” daquela atividade comercial.

Em geral, um dado isolado pode não ter a capacidade de provocar danos ao seu proprietário, no entanto, quando combinado com outros dados, pode causar sérios problemas, até mesmo destruir reputações. Esse é um dos focos dos chamados *data brokers* (corretores de dados): combinar vários dados cruzando distintas referências, para em seguida compartilhá-los e/ou comercializá-los.

Alguns *data brokers* são empresas de grande porte. A Palantir Technologies, por exemplo, é uma das *startups* de maior sucesso no Vale do Silício; fundada em 2003, vale atualmente cerca de 20 bilhões de dólares. Existem muitas outras, como Acxiom, CoreLogic, Datalogix, eBureau, que atuam “nos bastidores”, com as quais não interagimos diretamente, mas que monitoram e analisam continuamente nosso comportamento no ambiente digital.

Como adverte a matemática Hannah Fry, em *Hello World: Being Human in the Age of Algorithms* (Olá, mundo: sendo humano na era dos algoritmos), a prática prolifera em muitos países onde não existem as leis de proteção de dados.⁹⁴ Nos Estados Unidos, por exemplo, o governo de Donald Trump impôs um retrocesso: em março de 2017, o Senado votou pela eliminação das regras que impediam os *data brokers* de vender os dados de usuários sem o seu consentimento explícito. Essas regras haviam sido aprovadas em outubro de 2016 pela Comissão Federal de Comunicações, agência independente criada para regular as comunicações interestaduais por rádio, *wire*, televisão, satélite e cabo.

O dilema privacidade *versus* conveniência parece, de certa forma, superado. Estamos cada vez mais aceitando trocar nossos dados por serviços, desde que com benefícios explícitos – Google, Waze, Uber, Facebook e muitos outros. O paradoxo é que expressamos preocupação com a privacidade e, simultaneamente, adotamos massivamente esses serviços baseados na coleta e no uso de dados pessoais.

Benefícios e ameaças da automação inteligente do varejo

10.1.2020

Desde 2016, a varejista chinesa Alibaba investe pesado no conceito de *new retail* (novo varejo), termo cunhado pelo seu fundador, Jack Ma, para definir um ecossistema que combina os canais *online* e *offline*. No supermercado futurista do grupo, Hema Xiansheng, inaugurado em 2015, o pagamento é automático e processado por reconhecimento facial (tecnologia de inteligência artificial).

Seu competidor norte-americano, a Amazon, em 2018 inaugurou a mercearia Amazon Go, em Seattle, onde o cliente transita pela loja, põe os produtos na sacola e vai embora; um aplicativo baixado previamente no celular registra as compras e os pagamentos sem qualquer interferência humana. O foco de ambos é a conveniência do consumidor (sem caixas, sem filas, mais facilidade e agilidade).

As tecnologias digitais estão invadindo o varejo mundo afora, extrapolando as funções de meios de pagamento. A North Face – principal fornecedora mundial de vestuário, equipamentos e calçados – oferece um sistema de compras *online* interativo: o cliente, por meio do processamento de linguagem natural (ia), recebe recomendações personalizadas em conversa com o “vendedor”. Desenvolvido em parceria com a provedora de serviços Fluid e o Watson-ibm, as conversas geram dados que são transformados em *insights* para melhorar a experiência de consumo. O intuito é replicar nas plataformas eletrônicas a função do balconista de loja.

No Brasil, essas tecnologias estão sendo incorporadas no *e-commerce* e nas lojas físicas. Os mais visíveis são os totens de autoatendimento; o grupo gpa, por exemplo, já instalou 180 *self-checkouts* em 23 lojas das bandeiras Pão de Açúcar e Extra (20% do faturamento total dessas lojas). O Magazine Luiza, outro exemplo, inspirado nas lojas da Apple, transformou seus vendedores também em “caixas” (executam a operação financeira), sistema implantado em 100% das unidades Magalu. Segundo especialistas, essas facilidades tendem a reduzir em 30% o tempo do consumidor nas lojas. A automação do varejo, aparentemente, é uma tendência inexorável, exatamente por facilitar a vida do consumidor, que festeja as novidades.

Como nas demais implementações (saúde, educação, redes sociais, pesquisa, produção), as tecnologias inteligentes aportam inestimáveis benefícios e grandes desafios. No caso do varejo, o efeito negativo imediato é a eliminação de funções até então exercidas por trabalhadores humanos. Dos 16 milhões de postos de trabalho criados no Brasil entre 2003 e 2016, cerca de 9,2 milhões (70%) estão em risco, segundo estudo do Laboratório do Futuro, do Instituto Alberto Luiz

Coimbra de Pós-Graduação e Pesquisa de Engenharia da Universidade Federal do Rio de Janeiro (Coppe-ufrj). Independentemente dos números estimados, que variam conforme a metodologia da pesquisa, o consenso é de que as funções de baixa qualificação são as primeiras a ser substituídas pela automação inteligente (em geral, não por coincidência, são as mais numerosas e associadas ao “primeiro emprego”, o que tem impacto direto nos jovens).

Relatório do Fórum Econômico Mundial, de setembro de 2018, sobre o futuro do trabalho estima que até 2022 a mudança na divisão do trabalho entre humanos e máquinas/algoritmos afetará 75 milhões de cargos. Atualmente, nos setores cobertos pelo estudo do relatório, em média 71% do total de horas de tarefas são realizados por humanos, em comparação com 29% por máquinas/algoritmos; em 2022, essa média deve mudar para 58% de horas de tarefas executadas por seres humanos e 42% por máquinas/algoritmos.

As ameaças, contudo, não estão restritas ao potencial desemprego e ao aumento da desigualdade: as tão propagadas novas funções demandam certo nível de especialização e formação que a maior parte dos trabalhadores não possui, aumentando a competição pelas funções menos qualificadas, o que implica redução salarial e, conseqüentemente, de renda.

A base do funcionamento da economia do século xxi está nos dados (e não nos modelos/algoritmos de inteligência artificial), responsáveis pelo atual poder e concentração de mercado das *big techs* (as gigantes de tecnologia), e gradativamente estender esse poder a todas as organizações públicas e privadas com capacidade de gerar, armazenar e minerar grandes quantidades de dados para extrair informações valiosas sobre quase tudo. Do uso dos dados derivam questões como proteção à privacidade dos usuários, sofisticação dos mecanismos de persuasão (comunicação hipersegmentada e assertiva), controle excessivo por corporações e governos.

A complexidade desses modelos não recomenda análises e soluções simplistas baseadas em poucas variáveis ou em raciocínio dualista (bem e mal). Os modelos de inteligência artificial que estão permeando as atividades socioeconômicas aportam benefícios extraordinários e, simultaneamente, sérias ameaças. O desafio posto à sociedade é como conciliar ambos os efeitos.

O documentário *O dilema das redes* (2020), de Jeff Orlowski, lançado pela Netflix, é um manifesto-denúncia, inserido num movimento que convoca para a ação de mudar o modo como a tecnologia é projetada, regulamentada e usada, visando alinhá-la com “os interesses das pessoas, e não com os lucros”. O documentário trata de questões sensíveis, provocando debates acalorados em distintos fóruns, o que por si só justifica sua produção. A natureza viciante das mídias sociais é um problema a ser enfrentado pelos pais e educadores, particularmente dos adolescentes.

É mandatório conhecer a dinâmica dos mecanismos das redes sociais, mas é igualmente importante ponderar sobre alguns dos problemas do documentário: o protagonismo absoluto da tecnologia, quando inúmeras outras variáveis são corresponsáveis por cada uma das questões abordadas; a atribuição ao modelo de negócio das plataformas digitais a responsabilidade por males da sociedade contemporânea, inclusive alguns anteriores à internet e às redes sociais; e, provavelmente proposital, a parte ficcional exagera e deforma o funcionamento dos algoritmos de inteligência artificial (personalização não é sinônimo de individualização). Paira no ar a dicotomia entre o bem e o mal: *big techs* (vilões) *versus* outros setores e usuários (vítimas inocentes).

O modelo de negócio dessas plataformas é captar e extrair informações dos dados dos usuários e oferecer aos anunciantes a possibilidade de comunicar suas mensagens (anúncios publicitários) de forma hipersegmentada, a partir de um conhecimento relativamente mais profundo do comportamento de seus usuários. Esse é o elo comum entre o Google e o Facebook, ambos monetizam os dados da movimentação de seus usuários (no último trimestre de 2019, de 96,8% da receita do Facebook veio dos anúncios, e 3/4 destes são anúncios de pequenas e médias empresas). A maior parte das críticas do documentário é direcionada às redes sociais, o que não é o caso do Google. Não por acaso, o Facebook é atualmente a maior empresa de mídia, e sua plataforma móvel (celular) entrega os mais eficientes resultados do mercado (investimento de mídia *versus* retorno).

Os dados são a matéria-prima desses modelos de negócio; logo, o tempo de permanência e a intensidade da interação nas plataformas são elementos-chave: quanto maior a interação, mais dados são gerados e minerados, melhor a assertividade da hipersegmentação oferecida aos anunciantes, maior o faturamento. As grandes empresas de tecnologia são a parte mais visível da Economia de Dados, mas não a única; o uso do *big data* e de modelos preditivos de inteligência artificial, gradativamente, dissemina-se na sociedade.

Não existem evidências de manipulação com fins políticos ou ideológicos por parte das plataformas tecnológicas, aparentemente o que ocorreu nas eleições foi o uso do Facebook pelas campanhas como qualquer cliente ou anunciante: compra da hipersegmentação para direcionar a publicidade política. A solução, nesse caso, poderia ser regulamentar as campanhas eleitorais, proibindo as redes sociais de oferecer a hipersegmentação para finalidades eleitorais.

Yochai Benkler, Robert Faris e Hal Roberts analisaram exaustivamente o processo eleitoral norte-americano de 2016 no livro *Network Propaganda* (Propaganda em rede).⁹⁵ Com base em pesquisas empíricas, os autores inferem que a manipulação nas redes sociais não foi decisiva na eleição de Donald Trump, e que a polarização política decorre da dinâmica do ecossistema de mídia norte-americano. Para eles, o “culpado” dos males atuais não é a tecnologia, e os autores apontam o equívoco de governos, organizações da sociedade civil, acadêmicos e mídia que, ao tentarem entender o que está impulsionando a mudança global, atribuem o papel de vilão à tecnologia por ser o elemento novo no cenário. “A tecnologia nos permitiu analisar milhões de histórias publicadas em um período de três anos. A tecnologia nos permitiu analisar milhões de tweets e links e centenas de milhões de compartilhamentos e palavras do Facebook para dar sentido a essas histórias. E, no entanto, toda essa pesquisa habilitada pela tecnologia nos afastou da tecnologia como a principal variável explicativa de nossa atual crise epistêmica”, ponderam os autores.

O papel da tecnologia e da técnica é um dos temas mais debatidos na história da humanidade. A concepção que predominou ao longo de grande parte da história do Ocidente foi a visão instrumental da tecnologia de Aristóteles – qualquer técnica envolve uma criação, e a origem do que será gerado pelo exercício dessa técnica está localizado no fabricante, isto é, no homem, e não no produto, seres artificiais inferiores.

Dois filósofos do século xx, Gilbert Simondon e Martin Heidegger, ofereceram novas perspectivas, quebrando a supremacia interpretativa de Aristóteles. Em 1958, Simondon agrupou a relação com a técnica em duas atitudes culturais contraditórias: de um lado, os objetos técnicos são tratados como puro conjunto de materiais, destituído de qualquer significação, representando apenas um utilitário; e de outro, supõe-se que eles sejam animados de intenções hostis ao homem, representando uma permanente e perigosa ameaça de agressão, uma insurreição.⁹⁶ Trata-se da idolatria da máquina, que atribui a esses objetos uma existência separada, autônoma. A segunda perspectiva assemelha-se às “tecnorreligiões” apontadas por Yuval Harari em seu livro *Homo Deus*, particularmente o dataísmo, originado no Vale do Silício, em que os humanos gradativamente perdem relevância. Nem neutra nem determinista, no estágio atual a tecnologia é um elemento facilitador e alavancador de estratégias definidas pelos humanos. Algumas positivas, outras nem tanto.

Os mecanismos de persuasão, outro tema central do documentário, não são

novidade; a dimensão persuasiva da publicidade é pura expressão desses mecanismos. Os publicitários se cercam de profissionais especializados no entendimento do consumidor, em traçar o perfil do *target* visando criar uma comunicação capaz de interferir nas decisões de consumo, seja de informação, produto ou serviço. A propaganda por décadas influenciou e gerou novos hábitos, novos comportamentos, com impactos culturais tremendos. O objetivo de qualquer empresa capitalista é monetizar a interação com os seus consumidores. A propaganda foi essencial no desenvolvimento do capitalismo industrial, e a hipersegmentação é essencial no capitalismo de dados.

O campo da engenharia comportamental é anterior à publicidade, à internet, às redes sociais, aos algoritmos de ia. Burrhus Frederic Skinner, já nos anos 1930-1940, acreditava que os humanos poderiam ser condicionados como qualquer outro animal, e que a psicologia comportamental poderia e deveria ser usada para construir uma utopia tecnológica em que os cidadãos seriam treinados desde o nascimento. Shoshana Zuboff, presente no documentário, crê que essas plataformas tenham aperfeiçoado e complementado as ideias de Skinner.⁹⁷

O laboratório da Universidade de Stanford, Stanford Persuasive Tech Lab, onde o protagonista do documentário, Tristan Harris, estudou foi criado em 1998 por B. J. Fogg. Seu propósito é gerar *insights* para desenvolver tecnologias aptas a mudar as crenças, os pensamentos e os comportamentos dos indivíduos de maneira previsível. Os estudos incluem design, pesquisa, ética e a análise de produtos de computação interativa – computadores, celulares, websites, tecnologias sem fio, aplicativos móveis, videogames. Em artigo de 1998, Fogg define persuasão “como uma tentativa de moldar, reforçar ou mudar comportamentos, sentimentos ou pensamentos sobre um problema, objeto ou ação”.⁹⁸

O documentário cumpre a tarefa de alertar para os potenciais impactos negativos dos modelos de negócio baseados em dados, entre eles a concentração de mercado das *big techs*, mas é importante identificar o que mudou por conta do *big data* e dos modelos de inteligência artificial, e não esquecer que o “tempo livre” gerado com os avanços conquistados na segunda metade do século xx foi dedicado à televisão. Como lembra Clay Shirky, em *A cultura da participação*,⁹⁹ “desde a década de 1950, em todo o mundo desenvolvido, as três atividades mais comuns são trabalhar, dormir e ver tv”. Outro ponto de reflexão é que as tecnologias dessas plataformas entregam, em parte, o que os usuários querem receber, caso não fosse assim, não haveria uma adesão tão intensa e extensa.

Como afirma um dos entrevistados: “Vivemos a utopia e a distopia ao mesmo tempo”, daí advém a complexidade, por isso não há solução fácil. Nos países líderes no uso da ia, não por coincidência, até agora proliferam somente diretrizes e princípios gerais.

Cresce o desconforto com o controle dos novos espaços públicos pela esfera privada, particularmente pelas redes sociais. O mundo está mais sensível a questões como privacidade, mediação e personalização de conteúdo na internet. Os protestos do movimento Black Lives Matter, após o assassinato de George Floyd pela polícia de Minneapolis, nos Estados Unidos, deram visibilidade aos algoritmos discriminatórios dos sistemas de vigilância usados pelos departamentos de polícia, o que levou à suspensão temporária dos contratos com a Amazon, a Microsoft e a IBM. A pressão sobre as gigantes de tecnologia tem aumentado desde a eleição presidencial norte-americana de 2016, com as revelações da Cambridge Analytica.

O filósofo francês Pierre Lévy, em entrevista ao *Valor Econômico*, alerta que as *big techs* passaram a deter o monopólio da memória mundial, e alerta: “Elas estão desenhando uma nova forma de poder econômico, o que é evidente, mas sobretudo político. Muitas funções sociais e políticas, que são funções tradicionais dos Estados-nação, estão passando para essas companhias. Na minha avaliação, é uma nova forma de Estado, que eu denomino Estado-Plataforma”.¹⁰⁰

Em 20 de outubro, o Departamento de Justiça dos Estados Unidos e 11 estados norte-americanos entraram com uma ação contra o Google por concorrência desleal, alegando “operação casada” do seu mecanismo de busca com contratos de fabricantes de dispositivos móveis. Na visão da Justiça norte-americana, essa associação prejudica os consumidores ao privá-los da prerrogativa de escolha de concorrentes. O último caso de aplicação da lei antitruste norte-americana foi o julgamento da Microsoft – o processo iniciado em 1998 e encerrado em 2011 questionava a “venda casada” entre o sistema operacional Windows e o navegador Internet Explorer –, e o último desmembramento foi da at&t, em 1974.

O Google declarou em um *tweet* que a ação “está profundamente equivocada. As pessoas usam o Google por escolha própria, e não porque são obrigadas a fazê-lo ou porque não encontram alternativas”. O argumento é em parte legítimo, *stricto sensu* é fato que não somos obrigados a pesquisar no Google nem ter perfil no Facebook e no Instagram, nem nos comunicar via WhatsApp. Por outro lado, são inegáveis a eficiência e o alcance dos serviços oferecidos por essas plataformas. Na base de seus modelos de negócio estão os algoritmos de inteligência artificial, que são treinados e aperfeiçoados continuamente pelos dados gerados nas interações dos usuários. Essa dinâmica cria poderosas barreiras

de entrada de novos concorrentes.

A natureza do processo contra o Google Research é distinta da pressão sobre a suposta manipulação do Facebook, acusado, por exemplo, de censurar publicações favoráveis a determinados candidatos ou não impedir a publicação de *fake news*. O desafio é como e a quem cabe distinguir liberdade de expressão de atentado à democracia. Esse debate atinge o Google como dono do YouTube: desde fevereiro, por exemplo, foram publicados 200 mil vídeos enganosos sobre a covid-19.

Na Europa, a comissão de defesa de concorrência da União Europeia, Margrethe Vestager, pretende obrigar as empresas a abrirem seus arquivos de anúncios para reguladores e pesquisadores, prometendo anunciar, em 2 de dezembro, duas iniciativas para controlar as gigantes de tecnologia (*big techs*) – *Digital Services Act (Ato de Serviços Digitais)* e *Digital Markets Act (Ato de Mercados Digitais)*.

Proliferam sugestões de como reduzir o poder das *big techs*. Uma delas é os indivíduos serem proprietários de seus dados; nesse caso, as redes sociais seriam pagas, e os usuários, individual ou coletivamente, receberiam o aluguel dos anunciantes, podendo, inclusive, transferir os dados para outras redes. Glen Weyl, economista da Microsoft, propõe a formação de “sindicatos” para negociar em nome de grupos de usuários pelo direito a um percentual da receita gerada com o uso de seus dados (“dividendo digital”). Outra opção “na mesa de negociação”, sem proposta clara de viabilidade, é taxar os dados.

Josh Simons, pesquisador visitante na área de “Responsible AI” do Facebook, e Dipayan Ghosh, ex-consultor de privacidade e políticas públicas na mesma empresa (2015-2017), defendem que o Facebook e o Google devem ser regulamentados como serviços públicos.¹⁰¹ Para eles, sendo a esfera pública (*online* e *offline*) um espaço crítico de comunicação e organização, expressão política e tomada de decisões coletivas, ao controlar como essa infraestrutura é projetada e operada, essas plataformas concentram poder econômico, político e social (como denuncia, inclusive, a senadora americana Elizabeth Warren).

Essas ideias desarticulam os modelos de negócio das *big techs*; vale indagar se essa desarticulação atende aos interesses dos usuários, ou seja, qual o *trade off* entre privacidade/mediação/personalização/manipulação e benefícios. Qual seria o resultado de um plebiscito entre os usuários sobre se estariam de acordo com medidas severas contra essas plataformas, se comprometessem os benefícios ofertados?

O professor da Columbia Law School Tim Wu questiona se os níveis extremos da concentração atual do mercado são compatíveis com a premissa de igualdade entre os cidadãos, a liberdade econômica e a própria democracia.¹⁰² Para Wu, os legisladores precisam atualizar as regulamentações antitruste. Originada com a Lei Sherman (1890), nos Estados Unidos, a Defesa da Concorrência ou Antitruste foi promulgada em reação à formação de grandes

monopólios e cartéis no final do século xix. Contudo, foi só nas décadas de 1950-1960, pós-Segunda Guerra Mundial, que a lei antitruste foi claramente identificada como essencial num regime democrático. A natureza dos modelos de negócio dessas plataformas, contudo, ao oferecerem serviços “gratuitos” em troca de dados, dificulta identificar os efeitos negativos sobre seus usuários nos moldes dessas leis.

Em artigo no *Time* de 2018, o jornalista David Kirkpatrick relata uma conversa que teve com Mark Zuckerberg,¹⁰³ em que ele justifica a contratação de Sheryl Sandberg, em 2008, como chefe operacional do Facebook pela sua experiência anterior como chefe de gabinete de Larry Summers, quando este era secretário do Tesouro do presidente Bill Clinton. Zuckerberg teria dito em 2009: “Em muitos aspectos, o Facebook é mais como um governo do que como uma empresa tradicional. Estamos realmente definindo políticas”. A afirmação de Zuckerberg deve ser levada a sério. O problema é que, em ambientes complexos, as soluções não costumam ser simples.

Dados gerados pelos usuários nas plataformas digitais: bens comuns ou proprietários?

22.1.2021

Em 2011, movimentos sociais conhecidos como “Primavera Árabe” em poucos meses derrubaram governos (como na Tunísia e no Egito) ou terminaram em conflito sangrento (como na Líbia, no Iêmen e na Síria). Liderados por jovens antenados com as tecnologias digitais, eles romperam com o monopólio de mídia dos ditadores, desconstruindo suas mensagens e convocando a população árabe para agir em plataformas como Facebook, Twitter e YouTube.¹⁰⁴ No mesmo ano, irromperam movimentos sociais mundo afora, como Occupy Wall Street (ows), no Zuccotti Park, distrito financeiro de Nova York, e os Indignados, em diversas cidades da Espanha. O protagonismo das mídias sociais foi não só reconhecido, como também amplamente festejado.

Em anos recentes, contudo, cresce a hostilidade em relação às *big techs*, fenômeno cunhado pela revista *The Economist* como “*techlash*”: “Reação negativa forte e generalizada ao crescente poder e influência de grandes empresas de tecnologia, especialmente aquelas sediadas no Vale do Silício”, conforme já consta no *Oxford English Dictionary*. Essa hostilidade advém do poder desses conglomerados tanto na mediação da comunicação e da sociabilidade, quanto na economia e na política. As interferências das redes sociais nas eleições e a discriminação racial dos sistemas de vigilância, baseados em tecnologias de reconhecimento facial, foram alguns dos eventos que deram visibilidade a questões como a segurança, a privacidade e a propriedade dos dados pessoais, e as imperfeições dos modelos de inteligência artificial. Contribui, igualmente, a percepção de que seus produtos afetam a saúde mental, particularmente dos jovens adolescentes (natureza viciante, *cyberbullying*, conteúdos extremistas).

Essa hostilidade, portanto, não é monotemática e remete a várias reflexões. A primeira é sobre a opacidade dos algoritmos de inteligência artificial. A tecnologia de ia oferece à sociedade a oportunidade inédita de tornar mais transparentes os processos de decisão em vários domínios. As decisões humanas são absolutamente enviesadas, em função de crenças, repertórios, vivências de cada pessoa (o viés nos dados, utilizados para treinar os algoritmos de ia, reflete em parte os preconceitos humanos). É utópica a ideia de “transparência” no ambiente virtual, porque não o são as relações entre humanos, entre instituições e humanos, e entre as próprias instituições. Cada um revela o que deseja, de acordo com os seus interesses (além do inconsciente).

Outra reflexão é que o “mundo digital” não é um mundo à parte do “mundo físico” nem é um espaço homogêneo; contempla múltiplos atores: instituições, indivíduos, conexões, interesses, natureza, tecnologias, dispositivos, regulamentações. Ou seja, a sociedade é complexa, parte de suas atividades é presencial e parte é virtual, mas é o mesmo mundo (não obstante as especificidades), em que o principal ativo é a informação (pressuposto da Economia da Informação em Rede e da Economia de Dados).

Esses argumentos estão longe de serem consensuais, pelo contrário, estão no centro de intensos debates. Fernando Filgueiras e Virgílio Almeida defendem que uma das especificidades do “mundo digital” é o pressuposto de que os dados digitais são bens comuns.¹⁰⁵ A *Stanford Encyclopedia of Philosophy* refere-se a “bem comum” (*common good*) como as facilidades – materiais, culturais ou institucionais – que os membros de uma comunidade fornecem a todos os membros a fim de cumprir uma obrigação relacional que preserva os interesses comuns, e cita como exemplos canônicos os parques públicos, as estradas, a segurança pública e a defesa nacional, o sistema jurídico, os museus, o ar e a água limpos.¹⁰⁶

No mundo digital existem dados que são bens comuns, porque são gerados em plataformas compartilhadas – modelos de *peer production/wiki* (produção em pares, colaborativa) –, como aborda Yochai Benkler em diversos livros e artigos, a exemplo do seminal *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (*A riqueza das redes: como a produção social transforma os mercados e a liberdade*).¹⁰⁷ Benkler enfatiza o papel dos bens comuns de acesso aberto (*open source*) na inovação descentralizada e na pesquisa científica – pesquisa clínica de doenças raras, ciência cidadã, conjunto de dados sobre o oceano e atmosfera.¹⁰⁸

E existem os dados que não são bens comuns, porque são gerados em plataformas proprietárias, que é o caso dos dados gerados pelos usuários das redes sociais, como Facebook e YouTube: os acionistas dessas plataformas investem em larga escala para criar as suas “fábricas” (canais de comunicação, captação, processamento e armazenamento de dados, tecnologias, equipes), logo a matéria-prima produzida nesses ambientes, que são os dados digitais e constituem a base de seus modelos de negócio, são legitimamente proprietários.

Numa analogia simplista, o bem produzido por um trabalhador numa fábrica industrial é aceito, universalmente, como pertencente ao dono do capital, que, por sua vez, remunera o trabalhador pelos serviços prestados. No caso das plataformas digitais, o produto “dado digital” é gerado pelos usuários na sua forma bruta e se transforma efetivamente em dado digital “útil”, compartilhável, por conta da infraestrutura dessas plataformas proprietárias. Pode-se contestar que, nesse caso, o “trabalhador” não é remunerado, trabalha de graça para o dono da plataforma, o que não é verdadeiro: a “moeda” que remunera o usuário são os serviços ofertados pelas plataformas. Os serviços do Google e do Facebook, por

exemplo, não são gratuitos, pagamos pelo uso deles com os nossos dados (dizendo de outra forma, não entregamos nossos dados gratuitamente, somos remunerados com os serviços).

Um argumento contra essa analogia – fábricas industriais e plataformas digitais – é a relevância dessas plataformas na sociabilidade atual, consideradas por alguns como “espaço público”. O argumento é poderoso, mas não elimina as ponderações anteriores sobre a oportunidade de minimizar ou mesmo eliminar o viés dos processos de decisão por meio da inteligência artificial e de reconhecer que o mundo é único com atividades *offline* e *online*, e que os dados digitais podem ser bens comuns ou proprietários, a depender das condições em que foram gerados.

Tecnologias ambivalentes: conveniência *versus* pesadelo orwelliano

5.2.2021

Os assistentes virtuais, em geral, provocam fortes polarizações entre os que os consideram invasivos e os entusiastas dos seus benefícios. Aparentemente, os últimos estão “vencendo”: um em cada cinco adultos nos Estados Unidos tem um assistente de voz em casa, mercado dominado pela Alexa, da Amazon, com 70% de participação e mais de 100 milhões de dispositivos vendidos (seguida pelo Google Assistant, com 24%).¹⁰⁹ No Brasil, o produto está disponível desde 2019.

A Alexa interage com voz, reproduz música, faz lista de tarefas, define alarmes, transmite *podcasts*, reproduz audiolivros, fornece informações em tempo real sobre o tempo, trânsito, esportes, notícias, controla sistemas e aparelhos inteligentes e conectados. Todas essas interações com os usuários geram dados, que são coletados e armazenados, produzindo um conhecimento inédito de preferências e hábitos cotidianos. Como alerta Yuval Harari, os algoritmos estão nos observando, registrando aonde vamos, o que consumimos, com quem nos relacionamos, monitorando todos os nossos passos, respirações e batimentos cardíacos, conhecendo-nos melhor do que nós mesmos, podendo nos controlar e manipular. “Você pode ter ouvido falar que estamos vivendo na era de invadir computadores, mas isso dificilmente é a metade da verdade. Na verdade, estamos vivendo na era de *hackers* humanos”, ou seja, os humanos é que estão sendo invadidos.¹¹⁰

A tecnologia que processa as palavras e atende às solicitações são os algoritmos de inteligência artificial, particularmente a técnica de aprendizado profundo (*deep learning*). Com base nessa massa de dados, a técnica identifica correlações não visíveis aos humanos, aumentando fortemente o grau de assertividade das previsões sobre, por exemplo, o comportamento futuro dos usuários; no caso dos assistentes virtuais, eles selecionam a melhor resposta para as perguntas que lhes são feitas (personalização). Especialistas garantem que um simples “Bom dia” é suficiente para a Alexa detectar o “astral” do usuário, via o tom de voz e a respiração.

O tema da privacidade está no cerne dessa coleta e uso de dados sem precedentes. A ameaça à privacidade está presente também no monitoramento do processo por funcionários dos fabricantes dos dispositivos empenhados em detectar falhas e melhorar seus desempenhos (os funcionários escutam as conversas privadas, e, embora as gravações sejam anônimas, aparentemente, elas contêm informações suficientes para identificar o usuário); no processamento em nuvem, o que torna o sistema vulnerável a ataques de *hackers*; no

compartilhamento dos dados com desenvolvedores independentes, para criar novas funcionalidades, com construtores, para embutir os sistemas nos imóveis, e com fabricantes de aparelhos domésticos, com a mesma finalidade; além das imperfeições do próprio sistema, como o envio equivocado de mensagens.

O Halo Band, da Amazon, lançado em dezembro de 2020, é outro dispositivo controverso quando se trata de privacidade de dados pessoais. O Halo é uma pulseira *fitness* (ou de saúde, como preferem alguns) com sensor de temperatura e monitor de batimentos cardíacos, capaz de identificar se o usuário está aborrecido ou irritado pelo tom de voz. Para se diferenciar dos *smartwatches*, como o Apple Watch e o Fitbit, o dispositivo oferece habilidades inéditas que incluem visão computacional para varredura corporal (*scan*), processamento de linguagem natural para avaliar o tom de voz, além de análise do sono e rastreamento de atividades (recursos disponibilizados por meio de assinatura no valor de 3,99 dólares mensais). Segundo o médico líder do projeto, Dr. Malik Majmudar, o dispositivo tem uma abordagem holística da saúde (física e socioemocional).

Antecipando-se a futuros questionamentos sobre a privacidade dos dados, em 30 de outubro, antes do lançamento do produto, a Amazon publicou um *white paper* (relatório oficial com esclarecimentos sobre determinado tema) em que garantia que os dados de saúde coletados pelo Halo não são utilizados para recomendações de produtos ou publicidade, e não são vendidos.¹¹¹ Parece, contudo, não ter sido suficiente: em 11 de dezembro, a senadora norte-americana Amy Klobuchar enviou uma carta ao secretário do Departamento de Saúde e Serviços Humanos (hhs), Alex Azar, pedindo providências, principalmente porque o Halo não está sujeito à Lei de Responsabilidade e Portabilidade de Seguro Saúde (hipaa), de 1996, que estabelece padrões de segurança e privacidade para os dados médicos.

Essa “fome de dados” tem relação direta com os modelos de negócio das gigantes de tecnologia, que correlacionam eficiência e sociabilidade: quanto maior a sociabilidade (interações sociais), maior a geração de dados – o que implica aumento da eficiência de seus modelos. O empenho em identificar e mensurar preferências e hábitos dos usuários, e a partir daí prever comportamentos, é a lógica das plataformas e dos aplicativos tecnológicos, das redes sociais *online*, do comércio eletrônico e dos sites de busca como o Google.

A extração de informações valiosas dos dados, via algoritmos de inteligência artificial, permite direcionar a publicidade hipersegmentada, principal fonte de faturamento, por exemplo, do Facebook – no quarto trimestre de 2020, a receita de anúncios representou 96,8% do total –, além de otimizar produtos e serviços. Estabelece-se um círculo virtuoso: interações sociais geram dados; dados aprimoram os sistemas tecnológicos; e sistemas tecnológicos aprimoram as interações sociais.¹¹²

Reconhecendo que os arcabouços regulatórios apropriados demandam

tempo de maturação, como ação imediata poderia ser exigido das empresas de tecnologia lançar seus produtos acompanhados de um tipo de “bula”, em clara analogia à indústria farmacêutica. Diferentemente das configurações de privacidade, em geral escritas em letrinhas minúsculas e linguagem complexa, a bula explicitaria em formatos acessíveis as “contraindicações”, as situações em que os sistemas não devem ser adotados, e seus efeitos colaterais (riscos intrínsecos).

Lina Khan, presidente da ftc: ameaça vigorosa ao poder das gigantes de tecnologia

25.6.2021

Parte das maiores fortunas globais tem origem no varejo, como a de Samuel Walton, do Walmart; de Bernard Arnault, da lvmh, e de Amancio Ortega, da Zara (primeira e segunda maiores fortunas da Europa). Jeff Bezos, fundador da Amazon, atualmente é o “homem mais rico do planeta”. Na contramão da maior parte das empresas, a Amazon se beneficiou fortemente da covid-19: em 2020, obteve faturamento recorde (386,1 bilhões de dólares, contra 280 bilhões em 2019), dobrou os lucros trimestrais e conquistou a liderança em valor de marca.

Em junho, o presidente Joe Biden escolheu Lina Khan, professora da Faculdade de Direito da Universidade Columbia e crítica implacável do poder das gigantes de tecnologia, para presidir a Federal Trade Commission (ftc), órgão responsável por proteger os consumidores e a concorrência. Sua nomeação, a mais jovem a ocupar essa função (32 anos), é amplamente justificada; entre outras contribuições, destaca-se o artigo sobre a Amazon publicado no *Yale Law Journal*. Nele, Khan aborda com precisão a estratégia de negócio da Amazon e os equívocos dos reguladores norte-americanos.¹¹³

Responsável pelo seu extraordinário sucesso, a estratégia de negócio da Amazon baseou-se na disposição para perdas (sustentada pelos investidores) e na diversificação das linhas de negócio – desde 2015 a Amazon tem sido lucrativa, ampliando e fortalecendo a diversificação entre lojas *online* e *marketplace*, lojas físicas (mais de 500 lojas do Whole Foods Market), Amazon Prime e Amazon Ads. Negócios com margens elevadas, como Kindle (inclui o Kindle Direct Publishing), Echo/Alexa e Amazon aws, são líderes de categoria.

O programa de fidelização Amazon Prime, lançado em 2005, decisivo como impulsionador de crescimento (clientes prime gastam mais e são mais leais à plataforma), está presente em cerca de 70% dos lares norte-americanos de alta renda; em abril de 2021, o programa tinha mais de 200 milhões de membros em todo o mundo. Por anos, para ganhar *market share*, a Amazon Prime operou com prejuízo: estima-se que até 2011 cada assinante tenha gerado um prejuízo anual de 11 dólares (no Brasil, com menos benefícios, o programa está disponível desde 2019, em cerca de 90 cidades).

Em paralelo, a logística da Amazon tornou-se a base de grande parte do comércio eletrônico, gerando dependência, inclusive, dos concorrentes. Um fenômeno recente é a “competição cooperativa”: a Netflix, por exemplo, está entre os principais clientes da infraestrutura de nuvem da Amazon aws, concorrendo com o Amazon Prime em oferta de conteúdo.

A Amazon pauta-se pela obsessão da experiência do cliente, preços baixos, infraestrutura de tecnologia estável e geração de fluxo de caixa (coleta rapidamente os pagamentos dos clientes e paga os fornecedores com prazos relativamente longos, gerando liquidez para investir na expansão). A venda de produtos é responsável pela maior parte do faturamento da Amazon, mas, como tem custos elevados, opera com margens estreitas, amplamente compensadas pelas margens elevadas da venda de serviços, como assinatura Prime e aws.

Na primeira carta aos acionistas, em 1997, Bezos já explicitava sua estratégia de dominar o mercado sustentada por três pilares: base de clientes, valor de marca e infraestrutura operacional. Para entender o poder acumulado pela Amazon, Khan sugere considerá-la como uma entidade integrada, e não isolando suas diversas linhas de negócio e os preços praticados em cada segmento específico. Sem essa visão macro, não é possível capturar a real forma de dominação da empresa: alavanca as vantagens obtidas em um setor para impulsionar os demais.

Em sua crítica aos reguladores norte-americanos, Khan aponta o equívoco de julgar as práticas predatórias de preços da Amazon com base em premissas da velha economia: “O fato de a Amazon estar disposta a renunciar aos lucros pelo crescimento solapa uma premissa central da doutrina contemporânea de preços predatórios, que pressupõe que a predação é irracional precisamente porque as empresas priorizam os lucros em vez do crescimento”. Outro elemento não considerado pelos reguladores é o fato de a Amazon usar sua função de fornecedora de infraestrutura para beneficiar suas outras linhas de negócios. Alinhada com Tim Wu, Khan questiona a capacidade do atual arcabouço regulatório antitruste de enfrentar as práticas anticompetitivas das grandes plataformas de tecnologia. No livro *The Curse of Bigness*¹¹⁴ (*A maldição da grandeza*), Wu pondera que os níveis extremos de concentração de mercado são incompatíveis com a premissa de igualdade, com a liberdade industrial e com a própria democracia; e alerta: “Existe muito poder privado concentrado em poucas mãos, com muita influência sobre o governo e nossas vidas”. Para ele, uma nova lei antitruste com sistemas de controle atualizados sobre a concentração privada do poder econômico daria aos humanos uma chance de se proteger das corporações.

Em suas análises, tanto Lina Khan quanto Tim Wu não abordam o modelo de negócio *data-driven* das *big techs*. Originalmente criado pelo Google, a prática de extrair informações de grandes conjuntos de dados com técnicas de inteligência artificial e, com base nelas, fazer previsões tornou-se um fator crítico de sucesso das plataformas tecnológicas e barreira de entrada para novos concorrentes. O acesso a volumes extraordinários de dados (*big data*), gerados nas interações dos usuários em suas plataformas, forja um círculo virtuoso: a diversidade de atuação promove captação de dados igualmente diversificada (além de volumes inéditos), que aprimora o desempenho dos produtos e serviços, como o sistema pioneiro de recomendação da Amazon, atraindo mais usuários, que

geram mais dados.

A assistente virtual Alexa é um veículo poderoso de captação de dados qualificados sobre o comportamento cotidiano dos usuários. Em 2015, a Amazon expandiu seu alcance ao abrir para desenvolvedores terceirizados a possibilidade de agregar habilidades e, a partir de 2018, por meio de acordos com fabricantes de dispositivos domésticos, ao juntar ao comando da Alexa interruptores de luz, sistemas de segurança, fechaduras de porta, termostatos, campainhas, entre outros. Esses dados significam pesquisa massiva de preferências, a custo relativamente baixo.

Com acesso a recursos financeiros quase ilimitados, que permitiram investir na busca de escala (preços baixos para expandir a base de clientes), a Amazon tornou-se uma plataforma diversificada de negócios, em que as lojas *online* impulsionam outros negócios mais lucrativos, adquirindo um poder de influência que vem sendo considerado excessivo mundo afora. A aposta é alta na atuação de Lina Khan na ftc.

Desafio do capitalismo de dados: reaproximar economia e ética com base em novos paradigmas

15.10.2021

Em meados de 2013, com a revelação do esquema de espionagem da Agência de Segurança Nacional norte-americana (nsa) por Edward Snowden, o livro *1984*, de George Orwell (1949), teve um aumento de vendas de 7.000% na Amazon, passando da posição 13.074 para a 193 da lista de livros mais vendidos, e permanecendo entre os 100 mais vendidos entre 2013 e 2016. Em janeiro de 2017, uma semana após a posse de Donald Trump, o livro figurou como o mais vendido dentre todos os gêneros na Amazon. No contexto atual, a distópica Oceânia de Orwell é uma reflexão sobre tecnologia *versus* privacidade e controle, temas que estão na pauta da sociedade.

Em 5 de outubro, Frances Haugen, ex-gerente de produto do Facebook, testemunhou no Senado dos Estados Unidos sobre os danos das mídias sociais, particularmente, sobre a democracia e a saúde mental de adolescentes. O depoimento no Senado foi precedido de uma série de revelações publicadas pelo *Wall Street Journal*,¹¹⁵ com base em dezenas de milhares de documentos internos fornecidos por Haugen. Em sua denúncia, ela alega que os executivos estão cientes dos danos e sabem como evitá-los, mas não tomam providências, porque privilegiam os lucros, e não a segurança e a integridade das pessoas. Sendo verdade ou não, o teor do depoimento fortalece o movimento de regulamentação da inteligência artificial, tecnologia-chave dos modelos de negócio das plataformas de tecnologia (e dos processos de transformação digital das empresas tradicionais).

Mundo afora, proliferam propostas para reduzir o poder dessas plataformas, desde limitar o tamanho delas por país (estabelecendo um número máximo de usuários) até implementar “testes de segurança” supervisionados por reguladores antes de liberá-las ao público, aos moldes de outras indústrias, como a farmacêutica, a aeroespacial e a automotiva. São múltiplas as propostas de auditoria: pública ou privada, com transparência ou com confidencialidade, restrita ou não às aplicações de maior risco, com periodicidade aleatória ou regular.

Em termos de regulamentação, estão em debate a Proposta da Comissão Europeia, com 108 páginas,¹¹⁶ e o Projeto de Lei n.º 21/2020, com nove páginas, aprovado em 29 de setembro pela Câmara dos Deputados brasileira e enviado para apreciação e votação no Senado. Com escopos radicalmente distintos – a proposta europeia é extensa e rígida, e o projeto brasileiro é uma mera carta de princípios –, ambos não atendem a complexidade da tecnologia de inteligência

artificial.

Apesar de os algoritmos de ia serem a base dos modelos de negócio (*data-driven models*) do “capitalismo de dados”, conceituado por Viktor Mayer-Schönberger e Thomas Ramge como a nova configuração do capitalismo (as plataformas de tecnologia são sua parte mais visível, mas não a única),¹¹⁷ seus efeitos negativos permanecem abstratos para a maioria da sociedade, inclusive para os governos, reguladores e legisladores. Em 2019, por exemplo, os senadores Ron Wyden e Cory Booker, ambos do Partido Democrata, apresentaram ao Congresso dos Estados Unidos um projeto de lei sobre responsabilidade algorítmica – *Algorithmic Accountability Act*¹¹⁸ –, projeto que nem chegou a ser apreciado por uma das comissões do Senado, logo não foi submetido ao plenário.

Wyden e Booker pretendem reapresentar o projeto de lei, inclusive porque há sinais de que ele seja bem visto pelo governo do presidente Joe Biden (ao menos como referência inicial para uma futura legislação de inteligência artificial). O foco do projeto são as aplicações de alto risco, prevendo que as empresas com faturamento acima de 50 milhões de dólares (ou no controle de mais de 100 milhões de dados pessoais) sejam obrigadas a auditar seus modelos de inteligência artificial, aos moldes das avaliações de impacto ambiental. A partir da definição de “sistemas de decisão automatizada”, o projeto indica os critérios de avaliação de impacto desses sistemas (a publicação dos resultados é opcional). O órgão responsável pela fiscalização, com regras de intervenção previstas no projeto, é o Federal Trade Commission (ftc), atualmente presidido pela advogada Lina Khan.¹¹⁹

Em 1987, o economista e filósofo Amartya Sen, Prêmio Nobel de Economia (1998), publicou o livro *Sobre ética e economia*.¹²⁰ Nele, Sen sinaliza a conexão histórica entre economia e ética, conexão que remonta a Aristóteles (*Ética a Nicômaco*) e Adam Smith (considerado o “pai da economia moderna”, Smith era professor de Filosofia Moral na Universidade de Glasgow). A partir da década de 1930, a relação entre ética e economia foi minimizada; Sen defende, enfaticamente, a reaproximação. Ao correlacionar sociabilidade e eficiência – interações sociais geram dados, dados aprimoram os sistemas tecnológicos, gerando mais lucro –, o modelo de negócio *data-driven*, contudo, estabelece uma relação inédita entre ética e economia ainda longe de ser compreendida.

Em março de 2013, a contaminação do suco de maçã Ades colocou em xeque a reputação da marca Unilever, a confiabilidade em seus produtos e o reconhecimento de qualidade como um de seus principais atributos. Catorze pessoas entraram em contato com a empresa por meio do serviço de atendimento ao consumidor, alegando mal-estar pelo consumo do produto; desse total, 12 receberam atendimento médico e foram liberadas, e duas não quiseram ser atendidas. A multinacional, enfrentando sua pior crise em 84 anos de Brasil, teve de lidar com os órgãos reguladores (desde a Anvisa até a Defesa do Consumidor), atender as dúvidas dos consumidores (o sac recebeu cerca de 220 mil ligações) e

evitar potenciais processos legais.

No caso descrito, a ética está correlacionada com o interesse direto do consumidor, e a qualidade do produto é fator de sucesso comercial. O modelo de negócio das plataformas de tecnologia rompe com essa correlação tradicional: as mensagens de ódio, por exemplo, geram mais interação e propagam mais rapidamente, porque mobilizam mais os usuários, ou seja, aumentam a interação na plataforma (ou “consumo” do serviço ofertado pela plataforma), comprometendo a qualidade do serviço (qualidade como função do nível de dano causado ao usuário).

Compreender essa nova lógica demanda novos paradigmas. O termo “paradigma” deriva do grego *paradeigma*, significando “modelo” ou “padrão”, ou seja, um conjunto de regras e princípios, fruto do desenvolvimento cultural, histórico e civilizatório de uma época. Em pleno século xxi, o desafio é formular paradigmas compatíveis com o novo ambiente de negócios.



Iniciativas regulatórias

Em 25 de novembro de 2021, os 193 países membros da Unesco firmaram o primeiro acordo global sobre a ética da ia, *Global Agreement on the Ethics of Artificial Intelligence*.¹²¹ Com o propósito de garantir o desenvolvimento e o uso saudável da ia, o acordo estabelece uma estrutura normativa que atribui aos Estados a responsabilidade sobre a regulamentação e a fiscalização. Elaborado por um grupo multidisciplinar de 24 especialistas (Ad Hoc Expert Group, ahcg), o texto destaca as vantagens e os potenciais riscos da ia contemplando várias dimensões, entre elas o meio ambiente e as especificidades do Sul Global. No âmbito específico, o acordo enfatiza o controle sobre o uso de dados pessoais e proíbe, explicitamente, os sistemas de ia para pontuação social (aos moldes do sistema chinês de crédito social) e vigilância em massa.

Na última década, frente à disseminação da ia na sociedade, surgiram princípios gerais que, pela natureza abstrata, não são traduzíveis em linguagem matemática (vital para incorporar em modelos estatísticos), e tentativas de autorregulamentação ineficientes. No âmbito regulatório, com abordagens distintas, destacam-se a proposta da Comissão Europeia, *Artificial Intelligence Act* (aia), publicada em 21 de abril de 2021; o projeto de regulamentação dos algoritmos de ia do Cyberspace Administration of China (cac), em vigência desde 29 de setembro de 2021; e o Projeto de Lei n.º 21/2020 brasileiro, aprovado na Câmara dos Deputados em 29 de setembro de 2021 e remetido à avaliação do Senado Federal. Mundo afora, Estados, organizações de classe e organismos multilaterais lideram iniciativas regulatórias pontuais e/ou setoriais na busca de mitigar as externalidades negativas, particularmente das aplicações de ia de alto risco.

São bem-vindas as iniciativas de proteger a sociedade dos potenciais danos da ia e, particularmente, de seu uso pelas grandes empresas de tecnologia. Intervenções jurídicas e regulatórias não são suficientes, é crítico ampliar o conhecimento e a consciência da sociedade sem tecnofobia nem tecnoutopia.

Neste bloco, temos seis artigos que discorrem sobre desde as evidências da falha da autorregulamentação até os pontos de fragilidade das propostas de regulamentação. O segundo artigo pondera sobre o problema da lacuna de conhecimento, ou assimetria informacional, entre produtores de tecnologia e legisladores.

No livro *1984* (1949), George Orwell concebeu uma distopia chamada Oceânia em que a “realidade” é definida pelo governo. Seu personagem Winston Smith, um burocrata de meia-idade sem muito sucesso, trabalha no Ministério da Verdade, e sua função consiste em reescrever notícias do passado para adequá-las ao presente, perpetuando um regime político em que o controle social é exercido pela desinformação e pela vigilância constante.

Em 2013, durante o governo Obama, com a revelação do esquema de espionagem da Agência de Segurança Nacional (nsa), o livro de Orwell teve um aumento de vendas de 7.000% na Amazon; entre 2013 e 2016, ele permaneceu entre os 100 mais vendidos na Amazon e, em janeiro de 2017, uma semana após a posse do presidente Trump, figurou como o mais vendido dentre todos os gêneros. O repentino interesse explica-se pela sintonia com o ambiente atual, caracterizado pelas ameaças à privacidade, pela perda de relevância dos fatos com a valorização da versão e pelas novas formas de controle e de poder. Na atualidade, indivíduos, instituições e objetos estão conectados continuamente (a “internet das coisas”, ou em inglês *Internet of things* - IoT), e essa interação gera dados. Os dados representam o conhecimento acumulado sobre a sociedade e são protagonistas do denominado Capitalismo de Dados – expressão cunhada por Viktor Mayer-Schönberger e Thomas Ramge, autores de *Reinventing Capitalism in the Age of Big Data* (Reinventando o capitalismo na era dos dados) –, no qual as gigantes de tecnologia controlam parte da geração, da mineração e do uso dos dados.¹²²

Seus modelos de negócio são baseados na capacidade de identificar nos dados padrões, preferências e hábitos dos usuários e oferecer *insights* à tomada de decisão (valendo-se dos algoritmos de inteligência artificial, especificamente dos modelos de *machine learning/deep learning*). Essa lógica está nas plataformas e nos aplicativos tecnológicos, nas redes sociais, no comércio eletrônico e nos sites de busca como o Google, cujos designs são concebidos para ampliar a permanência de seus usuários, gerar engajamento e, consequentemente, produzir mais dados.

A complexidade torna esses modelos relativamente opacos e temerários, motivando o conceito de “Capitalismo de Vigilância” – proposto por Shoshana Zuboff em *A era do capitalismo de vigilância* –, no qual os lucros derivam da vigilância unilateral e da modificação do comportamento humano para fins de monetização. Para Zuboff, os dados sobre onde estamos, para onde vamos, como estamos nos sentindo, o que estamos dizendo, os detalhes de nossa direção e as condições de nosso veículo estão se transformando em receita na atual perspectiva

comercial.

A nova ordem econômica reivindica a experiência humana como matéria-prima gratuita para práticas comerciais ocultas de extração, previsão e vendas. Em sua visão, o novo capitalismo, em vez de produzir produtos, gera lucro extraindo, analisando e vendendo dados (as plataformas tecnológicas, em geral, não vendem os dados, vendem publicidade direcionada/hipersegmentada).

Sem deixar de reconhecer os enormes benefícios das novas tecnologias, Brett Frischmann, Villanova University e de Stanford, e Evan Selinger, Rochester Institute of Technology, autores de *Re-Engineering Humanity*, argumentam que o recurso de coleta de dados faculta a governos e empresas um inédito poder de vigilância e a prerrogativa de usar as informações específicas coletadas e agir a partir delas.¹²³ Para os autores, as tecnologias estão desempenhando um papel essencial na formação de nossas crenças, emoções e bem-estar e, em geral, por mecanismos que entendemos parcialmente.

Os algoritmos de ia não apenas são instrumentos comerciais, mas também permitem prever e interferir, de maneira inédita, em nossa conduta em todas as esferas da vida social. O grau e eficácia dessa intervenção, contudo, é controverso.

O campo de pesquisa das tecnologias persuasivas vem se disseminando com o avanço da ia, com distintos propósitos, desde estimular vendas em estratégias comerciais até interferir nas escolhas políticas nos processos eleitorais. O cientista comportamental B. J. Fogg, por exemplo, fundou o Stanford Persuasive Tech Lab, em 1998, com o intuito de gerar *insights* no desenvolvimento de tecnologias aptas a mudar as crenças, os pensamentos e os comportamentos dos indivíduos de maneira previsível. Denominado de *captology*, o estudo das tecnologias persuasivas inclui design, pesquisa, ética e a análise de produtos de computação interativa. Seus pesquisadores conceituam a “captologia” como uma nova maneira de pensar sobre o comportamento-alvo e transformá-lo numa direção compatível com o “problema” a ser resolvido.

Parte dos riscos está sendo mitigada em leis de proteção de dados, como a Lei Geral de Proteção de Dados do Brasil (lgpd). Diretamente sobre as tecnologias de inteligência artificial, tramita no Senado o Projeto de Lei n.º 5.051/2019, do senador Styvenson Valentim (Podemos-RN), que contém os princípios que o desenvolvimento da ia deveria seguir e restringe o papel dos sistemas de IA ao de mero apoiador da tomada de decisão humana. Seus críticos argumentam que o baixo número de artigos (apenas sete) deixa de contemplar muitas questões relacionadas ao uso da tecnologia; outro questionamento é a ausência de descrição detalhada do termo “inteligência artificial”.

As iniciativas regulatórias ainda são tímidas (algumas equivocadas) em comparação com a dimensão dos desafios, mas sinalizam as primeiras reações positivas da sociedade.

Decodificando o cérebro humano: assimetria informacional entre produtores de tecnologia e legisladores

5.3.2021

A possibilidade de transferência da mente humana para computadores é um tema do universo da ficção científica. Em *Transcendence* (2014), filme de Wally Pfister, por exemplo, o pesquisador de inteligência artificial Dr. Will Caster, personagem interpretado por Johnny Depp, vive a experiência de ter sua consciência transferida para uma máquina. Se na ficção causa desconforto momentâneo, a mera possibilidade de concretização na “vida real” levanta inúmeras questões éticas que, aparentemente, não estão na pauta de prioridades dos formuladores de políticas e arcabouços regulatórios mundo afora. Ainda estamos longe de equacionar impactos mais visíveis e imediatos da inteligência artificial, como a mediação da comunicação e da sociabilidade e os processos decisórios automatizados.

Jack Gallant, professor de Psicologia e Neurociência da Universidade da Califórnia, tem como foco de pesquisa a modelagem computacional do cérebro humano. Para tal, desenvolve modelos (algoritmos de decodificação do cérebro) para descrever com precisão como o cérebro representa o mundo, como codifica as informações das atividades cotidianas, complexas e simples, com base em dados coletados por meio de ressonância magnética. Apoiado em “engenharia reversa” do algoritmo de reconhecimento de imagem (aprendizado de máquina/IA), Gallant e equipe buscaram visualizar o que uma pessoa estava vendo com base em sua atividade cerebral. Os primeiros resultados científicos, publicados em 2011, sinalizaram a possibilidade de, no futuro, uma máquina ler pensamentos e acessar memórias de humanos.

Volumes imensos de recursos estão sendo investidos na direção de transformar radicalmente a interação homem-máquina. O jornalista Moises Velasquez-Manoff, num longo artigo publicado no *New York Times*, descreve, entre outras pesquisas desenvolvidas por diferentes instituições, o estudo, financiado pelo Facebook, de Edward Chang, neurocirurgião da Universidade da Califórnia: produzir um capacete de leitura cerebral usando luz infravermelha. Recentemente, um de seus experimentos previu, com 97% de precisão, cerca de 250 palavras num conjunto de 50 sentenças utilizadas por um voluntário (aparentemente, a melhor taxa alcançada em estudos semelhantes).¹²⁴

Elon Musk e mais oito investidores criaram em 2016 a Neuralink, empresa de neurotecnologia com foco no desenvolvimento de interfaces implantáveis cérebro-máquina (*brain machine interface*, bmi). O propósito é transferir a mente humana para um computador, libertando o cérebro do corpo, num processo

denominado “*mind-upload*” (transferência da mente humana). A *startup* recebeu 158 milhões de dólares em financiamento, sendo 100 milhões de Musk (em julho 2019, contava com 90 funcionários).

Em *Superinteligência*, livro de referência sobre o futuro da inteligência artificial, o filósofo inglês Nick Bostrom pondera que a interface direta cérebro-computador permitiria aos humanos explorar os pontos fortes da computação digital – recordação perfeita, cálculo aritmético rápido e preciso, e transmissão de dados de alta largura de banda.¹²⁵ Outra possibilidade aventada por Bostrom é o aprimoramento gradual de redes que liguem as mentes humanas umas às outras, o que ele denominou de “coletivo superinteligente”.

As barreiras à concretização dessas ideias são imensas. Para o neurocientista da Universidade Federal do Rio de Janeiro (ufrj) Roberto Lent, não é possível visualizar o *upload* do cérebro para uma máquina num horizonte razoável, pelo gigantismo da circuitaria cerebral: “Como transferir 4 trilhões de sinapses para *bits*?”. Por outro lado, Lent admite que já existem mecanismos para estudar a interação remota entre duas ou mais pessoas por meio de técnicas de neuroimagem funcional. “Quando tem liga entre as atividades cerebrais, posso descobrir quais regiões dos cérebros estão sincronizadas, significando que posso me comunicar com você sem te ver, sem te ouvir, posso ter acesso no meu cérebro à sua atividade neural.” Segundo ele, é possível visualizar a comunicação direta entre cérebros num horizonte relativamente próximo.¹²⁶

Com os instrumentos atuais, contudo, é difícil estudar o cérebro humano: a pesquisadora Esther Landhuis, em 2017, comparou o tempo de computação e análise utilizado por pesquisadores chineses para estudar as conexões dos 135 mil neurônios de uma mosca – cerca de 10 anos – com o tempo necessário para estudar o cérebro humano, com seus 86 bilhões de neurônios, e concluiu que seriam necessários 17 milhões de anos.¹²⁷

Os enormes benefícios das tecnologias de inteligência artificial convivem com externalidades negativas a serem enfrentadas. Proliferam propostas de governança da ia, como a “auditoria baseada na ética” defendida pelo filósofo italiano Luciano Floridi e diversos outros pesquisadores de empresas como Google e Intel, e universidades como Oxford, Cambridge e Stanford. Para ser eficaz, ou seja, confiável, como reconhece Floridi, o auditor teria de ser um órgão governamental ou associado a uma organização multilateral. As propostas, em geral, esbarram na complexidade da tecnologia *versus* a lacuna de conhecimento dos legisladores e reguladores.

Em artigo publicado no site do Fórum Econômico Mundial (WEF), os pesquisadores Adriana Bora e David Alexandru Timis admitem a falta de clareza dos reguladores sobre as funções dessas tecnologias, ilustrando com as audiências no Congresso dos Estados Unidos de testemunhos de executivos de gigantes de tecnologia. “Ofereceram ao público a oportunidade de observar a preocupante lacuna de alfabetização digital entre as empresas que produzem a tecnologia que

está moldando nossas vidas e os legisladores responsáveis por regulamentá-la”, ponderam os autores.¹²⁸

A tendência é essa assimetria informacional aumentar na medida em que aumenta a complexidade dos modelos de inteligência artificial. Se aplicações básicas, como os algoritmos de seleção e classificação de conteúdo utilizados pelas redes sociais, não são plenamente acessíveis, o que dirá dos algoritmos de decodificação do cérebro. O desafio é evitar que a legislação chegue tarde demais, quando a tecnologia já estiver incorporada na sociedade.

Proposta europeia de regulamentação da inteligência artificial: impressões preliminares

30.4.2021

As tecnologias não são todas iguais, algumas adicionam valor incremental à sociedade, e outras são disruptivas. Ao reconfigurar a lógica de funcionamento da economia e aportar modelos inéditos de negócio, as disruptivas provocam períodos de reorganização, o que o economista Joseph Schumpeter denominou de “destruição criativa”. As tecnologias de uso geral (*general purpose technologies*, gpt) estão nesse último bloco. São tecnologias-chave, moldam toda uma era e reorientam as inovações nos setores de aplicação, como a máquina a vapor, a eletricidade e o computador. A inteligência artificial é a tecnologia de propósito geral do século xxi. Esse pressuposto, ou natureza, sugere regulamentações setoriais, incorporando as especificidades da ia aos arcabouços jurídicos e órgãos fiscalizadores preexistentes. Contudo, esse não foi o caminho adotado pela Comissão Europeia.

Na tentativa de proteger os humanos antes que seja “tarde demais”, a Comissão Europeia, órgão executivo da União Europeia, divulgou sua proposta de regulamentação do desenvolvimento, da implantação e do uso da ia (excetuando as aplicações para fins militares), fruto de processo iniciado em 2018. Ao longo de 108 páginas estão previstos procedimentos para fornecedores e usuários, com muitas significativas para situações de não conformidade. Não por acaso, o texto é permeado de lacunas, ambiguidades, excesso de adjetivos, imprecisões, em geral, não compatíveis com uma proposta de regulamentação.¹²⁹

Apesar da expressa intenção de estimular o desenvolvimento e a implementação da inteligência artificial na Europa, a rigidez das regras e o custo associado para atendê-las apontam para a direção oposta e, provavelmente, serão contestados ao longo do debate a que a proposta será submetida antes de se transformar efetivamente em regulamentação. Pelas reações iniciais, o rol de contrariados é extenso; além, obviamente, do setor privado, incluem-se o setor público, para o qual a ia tem sido útil para lidar com grandes conjuntos de dados (como no caso da Justiça brasileira), e instituições da sociedade civil que, na direção oposta, temem que a excessiva amplitude fomenta a ideia de autorregulação.

A proposta de regulamentação europeia é mais um tema para o governo Biden, já envolvido com a regulamentação das *big techs*. A expectativa é que o crescente protagonismo da China incentive o diálogo entre Estados Unidos e Europa, que terá de contemplar, entre outras, as diferenças entre seus sistemas jurídicos (*common law* norte-americano, alicerçado nos costumes e na

jurisprudência, e *civil law* europeu, fundamentado em leis e direito positivado).

A abordagem da proposta é baseada em risco, delimitando os sistemas e seus usos em três categorias: “*unacceptable risk*” (risco inaceitável), “*high risk*” (risco elevado) e “*low/minimal risk*” (risco baixo/mínimo). As aplicações de “*unacceptable risk*”, por representarem uma ameaça à segurança e aos indivíduos, serão proibidas: a) sistemas com técnicas subliminares para distorcer o comportamento de um usuário de maneira a causar danos físicos ou psicológicos, ou manipular o comportamento humano; b) sistemas que explorem vulnerabilidades de um grupo específico de usuários relacionado a idade, deficiência física ou mental, para distorcer materialmente o comportamento de forma que cause danos físicos ou psicológicos; c) sistemas utilizados por autoridades públicas para avaliar ou classificar a confiabilidade de pessoas físicas com pontuação social, ocasionando tratamento desfavorável. Igualmente, será proibida a identificação biométrica remota em “tempo real” em espaços acessíveis ao público (por exemplo, reconhecimento facial nas câmeras de vigilância), com poucas exceções – busca por criança desaparecida, ameaça terrorista ou para localizar um suspeito de crime grave –, e mesmo nesses casos o uso sujeita-se à prévia autorização judicial.

As aplicações de “*high risk*” serão rigorosamente monitoradas e referem-se ao uso de inteligência artificial em: a) infraestrutura crítica, por exemplo, transporte, que coloque em risco a vida e a saúde dos cidadãos; b) educação ou formação profissional, que determine o acesso como classificação de exames e seleção; c) componentes de segurança de produtos, como em cirurgias assistidas por robôs; d) gestão de trabalhadores, como *software* de análise de currículos em processos de recrutamento; e) serviços públicos e privados essenciais, como pontuação de crédito para obtenção de empréstimo; f) aplicação coerciva da lei, como avaliação da fiabilidade de provas; g) gestão de migração, de asilo e controle de fronteiras, como verificação da autenticidade de documentos de viagem; e h) administração da justiça e processos democráticos (aplicação da lei baseada em fatores específicos).

Nessa categoria, o monitoramento prevê instrumentos adequados de avaliação e mitigação de risco; registro das atividades para garantir a rastreabilidade dos resultados; documentação pormenorizada, permitindo a fiscalização das autoridades; supervisão humana em todas as etapas; e elevado nível de solidez, segurança e exatidão. Está previsto, igualmente, controle sobre os dados, desde a coleta até a preparação (rotulagem, limpeza, agregação), incluindo a obrigatoriedade de exames para detectar potencial viés (discriminação) durante toda a “vida útil” do sistema.

Ademais, os sistemas de “*high risk*” devem garantir que o seu funcionamento seja suficientemente transparente aos usuários, com instruções de uso em formato digital (“informações concisas, completas, corretas e claras que sejam relevantes, acessíveis e compreensíveis para os usuários”). Os sistemas

também devem permitir aos supervisores interpretar os resultados, compreender plenamente suas capacidades e limitações, e monitorar, detectar e resolver as anomalias, as disfunções e o desempenho inesperado. A proposta prevê um botão de “parar” a ser acionado pelo supervisor. Além disso, deve-se garantir que nenhuma ação ou decisão seja tomada pelo usuário com base na indicação do sistema, sendo obrigatória a confirmação de pelo menos duas “*singular person*” (sem definição explícita).

A proposta afirma que a maior parte das aplicações está na categoria “*low/minimal risk*”, mas a realidade indica o contrário: parte significativa das aplicações atuais, isoladas ou inseridas em modelos de negócio, está na categoria “*high risk*”. São muitas as “áreas cinzentas”; por exemplo, os aplicativos de inteligência artificial “manipulativos”, que visam alterar e influenciar o comportamento dos usuários, incluem ou não os algoritmos de ia dos modelos de negócio das plataformas digitais (instrumentos de persuasão e mídia hipersegmentada, como Google e Facebook)? Como restringir o uso de identificação biométrica em sistemas de vigilância aos casos previstos, se a efetividade depende de um sistema previamente instalado e ativo?

Na ponderação sobre regulamentação macro *versus* setorial, cabe refletir: se cada país tem seu Banco Central, que regula todo o funcionamento do sistema financeiro, qual o sentido de outro órgão definir e fiscalizar os procedimentos de concessão de crédito com inteligência artificial? Se cada país tem seu Ministério da Educação, qual o sentido de outro órgão definir e fiscalizar os procedimentos de aplicação da ia na educação? Se cada país tem sua Justiça, qual o sentido de outro órgão definir e fiscalizar quais modelos de ia podem ou não ser usados nos processos jurídicos? Se cada país tem seu órgão regulador da saúde, qual o sentido de outro órgão definir e fiscalizar a conduta médica associada à ia? Se cada universidade tem seu Comitê de Ética, qual o sentido de um órgão externo apreciar os projetos com ia? Ademais, quem seriam os membros desse super “comitê central” cuja missão é fiscalizar a partir de conhecimento tão especializado e critérios tão subjetivos? Ao optar por regulamentação macro, a Comissão Europeia enfrentará desafios que serão equacionados, ou não, no que se avizinha ser um longo percurso.

Inteligência artificial e as idiossincrasias do poder público brasileiro

9.7.2021

O Palasnet é um sistema usado pela Polícia Federal com base em técnicas de reconhecimento facial, ou seja, inteligência artificial. Existem outros exemplos de uso de ia pela polícia brasileira, como o sistema NuDetective, para detecção automática de nudez e pornografia infantil na internet. O Poder Judiciário, gradativamente, tem adotado sistemas de inteligência artificial. Pesquisa realizada pela Fundação Getulio Vargas (fgv), entre fevereiro e agosto de 2020, identificou 64 projetos em funcionamento ou em fase de implantação em 47 tribunais do país.¹³⁰

No âmbito do Supremo Tribunal Federal (stf) destaca-se o Projeto Victor. Desenvolvido em parceria com a Universidade de Brasília, é o primeiro projeto de inteligência artificial aplicado a uma Corte Institucional; seu objetivo é analisar e classificar, por temas recorrentes, os recursos extraordinários.¹³¹ Também existem iniciativas locais, como a plataforma Sinapses, do Tribunal de Justiça de Rondônia, com o intuito de estimular o desenvolvimento de modelos de ia adequados ao Processo Judicial Eletrônico (pje). Em agosto de 2020, o Conselho Nacional de Justiça (cnj) promulgou a Resolução n.º 332, que dispõe sobre princípios éticos a serem observados no desenvolvimento e no uso da inteligência artificial pelo Judiciário.¹³²

Na última década, em face da proliferação das tecnologias de ia na automação de distintas tarefas, um conjunto de princípios gerais vem sendo replicado mundo afora. Originado na Asilomar Conference on Beneficial ai, realizada em 2017 pelo Future of Life Institute, são conhecidos como “Asilomar Principles”.¹³³ Esses princípios gerais são de aplicabilidade restrita, ou seja, não são traduzíveis em boas práticas para nortear o ecossistema de inteligência artificial. Alguns deles, como justiça e dignidade, não são universais, e, mais desafiador, desconhece-se como representá-los em termos matemáticos, condição para que possam ser incorporados em modelos estatísticos de probabilidade.

A única proposta de regulamentação conhecida é a da Comissão Europeia.¹³⁴ A proposta é fruto de um processo relativamente longo de debate na comunidade europeia, envolvendo especialistas da academia, do mercado e do governo, que resultou em relatórios densos submetidos à apreciação pública. Com todo esse antecedente, mesmo assim, a proposta é vaga, confusa, em certos aspectos idealista, e a expectativa é de que seja debatida no decurso dos próximos anos antes de se transformar, se for o caso, efetivamente em lei.

No Brasil, com significativo atraso em relação aos países desenvolvidos, a inteligência artificial está em uso não só no setor público – principalmente na Justiça –, como também no segmento mais avançado do setor privado (principalmente nas plataformas e aplicativos que os brasileiros acessam no cotidiano, em sua quase totalidade desenvolvidos e controlados por empresas norte-americanas). Pelos seus benefícios – redução de custo e aumento de eficiência operacional, experiências inéditas para usuários, clientes e consumidores, previsões mais assertivas, decisões mais consistentes e uniformes –, a tendência é de adoção crescente, o que justifica amplamente iniciativas governamentais para proteger os cidadãos e as instituições dos potenciais impactos negativos. Mas não justifica a precipitação, responsável por gerar soluções incompletas, equivocadas e sem eficácia.

Em 9 de abril de 2021, o Ministério da Ciência, Tecnologia e Inovações (mcti) publicou no *Diário Oficial da União* a “Estratégia Brasileira de Inteligência Artificial” (ebia) com o propósito de: “Nortear o Governo Federal no desenvolvimento das ações, em suas várias vertentes, que estimulem a pesquisa, inovação e desenvolvimento de soluções em inteligência artificial, bem como, seu uso consciente, ético e em prol de um futuro melhor”. Comparando com as oito principais estratégias nacionais – de Estados Unidos, Canadá, Reino Unido, China, Índia, França, Alemanha e Coreia –, conclui-se facilmente que o documento do mcti é uma “não estratégia de ia”: carece de objetivos, metas, orçamento, cronograma, enfim, de todos os elementos que compõem um plano estratégico.

No dia 6 de julho de 2021, a Câmara dos Deputados aprovou o regime de urgência para o Projeto de Lei n.º 21/2020, de regulamentação da inteligência artificial, com autoria do deputado Eduardo Bismarck (pdt-ce), prevendo a votação da proposta nas próximas sessões do Plenário. Com sete páginas de texto (espaçadas), assim como a suposta estratégia brasileira de ia não é uma estratégia, o projeto de lei é um “não projeto de lei”, se tomarmos como referência a única proposta similar, a já citada aia, da Comissão Europeia (com 108 páginas). Entre inúmeras lacunas, o projeto não prevê punições nem multas, o que o torna inócuo do ponto de vista legal; não distingue claramente desenvolvedores de usuários; atrela a segurança a padrões internacionais de ia, sem atentar para o fato de que não existem tais padrões internacionais (além disso, uma lei tem de ser explícita para que sirva como base posterior de processos legais); define vagamente a prestação de contas pelos “agentes de inteligência artificial”; declara funções a serem exercidas pelo “poder público”, sem indicar o órgão responsável; prevê que União, estados e municípios recomendem padrões e boas práticas, sem especificar quais seriam. Nomeada de “consulta pública”, alguns especialistas em ia foram convocados pela Câmara dos Deputados para comentar o projeto em “extensos” 10 minutos!

Um projeto de lei confere direitos que serão invocados em situações de

arbitragem, logo é mandatório um conteúdo suficientemente claro, preciso e incondicional. Qualquer projeto dessa natureza precisa ser debatido intensamente pela sociedade para gerar conhecimento e consciência, inclusive para capacitar os deputados de modo que possam exercer suas atribuições de legislar com conhecimento de causa. Fica a pergunta: por que a urgência? O país, em plena crise sanitária-econômica-política-social, certamente, tem outras prioridades.

A China tem ética para o uso de dados e inteligência artificial, mas o poder decisório irrestrito é do governo

23.7.2021

A Administração do Ciberespaço da China (cac) ordenou que os aplicativos da Didi Chuxing – 377 milhões de usuários e 13 milhões de motoristas ativos, dona do aplicativo 99 no Brasil – fossem retirados das lojas por violação grave das leis de coleta e uso de informações pessoais. A suspensão ocorreu quatro dias após a abertura de capital na Bolsa de Valores de Nova York. A medida representa um marco no esforço das autoridades chinesas em definir normas e padrões éticos com relação à privacidade, inclusive mais rígidos do que da gdpr, lei de proteção de dados europeia.

Na prática, contudo, o poder decisório irrestrito do governo chinês relativiza a eficácia das medidas. O sistema jurídico na China está sujeito à supervisão e à interferência do Poder Legislativo, ou seja, do Partido Comunista. Isenções problemáticas emergem quando relacionadas, particularmente, à segurança, à saúde e ao “interesse público significativo”. Ao mesmo tempo que o governo estimula a coleta de grandes volumes de dados para atender o sistema de crédito social (*Social Credit System*), sem respeitar a privacidade, pressiona as empresas de tecnologia para que se enquadrem na legislação (enquadramento relativizado em função de interesses do governo).

O tratamento dos dados ganhou relevância na China com o *New Generation of Artificial Intelligence Plan* – AIDP¹³⁵ (*Plano de Desenvolvimento de Inteligência Artificial de Nova Geração*) primeiro esforço legislativo em nível nacional com foco em ia. As tecnologias de inteligência artificial já estavam presentes nos planos econômicos anteriores, mas a vitória no jogo de tabuleiro *Go*, em março de 2016, do sistema de ia AlphaGo da DeepMind/Google sobre o sul-coreano Lee Sedol, campeão mundial, acompanhada ao vivo por mais de 280 milhões de chineses, ensejou o reconhecimento do papel estratégico da ia no desenvolvimento econômico. Em 2017, o presidente Xi Jinping anunciou o novo plano aidp para uma audiência de diplomatas estrangeiros, declarando que a China seria líder mundial em ia até 2030.¹³⁶

O aidp tem três áreas de concentração: concorrência internacional (até 2020, manter a competitividade com outras grandes potências, otimizando seu ecossistema de inteligência artificial); crescimento econômico (até 2025, alcançar um “grande avanço” teórico no campo da ia e ser líder mundial em algumas aplicações); e governança social (ser o centro de inovação mundial, atualizando leis e normas para lidar com os novos desafios). A coordenação do plano é compartilhada entre o Ministério da Ciência e Tecnologia e o Escritório de

Promoção do Plano de Inteligência Artificial, assessorados pelo Comitê Consultivo de Estratégia de Inteligência Artificial, fundado em novembro de 2017. A ideia é que o aidp seja um incentivador ativo de projetos locais (e não centralizador).

O plano aidp elegeu algumas empresas privadas como “campeãs nacionais de ia”, com a função de desenvolver setores específicos: a Baidu, direção autônoma; a Alibaba, cidades inteligentes; e a Tencent, visão computacional para diagnósticos médicos. Os termos do acordo estabelecem que a empresa privada adotará os objetivos estratégicos do governo em troca de contratos preferenciais, acesso facilitado a financiamento e proteção de participação no mercado.

Em março de 2019, o Ministério da Ciência e Tecnologia da China criou o Comitê Nacional de Especialistas em Governança de Inteligência Artificial, que lançou princípios éticos básicos para o desenvolvimento da ia: foco na melhoria do bem-estar comum da humanidade, respeito pelos direitos humanos, privacidade e justiça, transparência, responsabilidade, colaboração e agilidade. Em paralelo, a Administração de Padronização da República Popular da China, órgão responsável pelo desenvolvimento de padrões técnicos, também lançou um documento com a defesa de um conjunto de princípios éticos e a proteção da propriedade intelectual.

Alinhados com as iniciativas governamentais, órgãos afiliados ao governo e empresas privadas criaram seus próprios princípios éticos de inteligência artificial. A Academia de Inteligência Artificial de Pequim, por exemplo, órgão de pesquisa e desenvolvimento de empresas e universidades, divulgou os “Princípios da ia de Pequim”: fazer o bem para a humanidade, usar a ia “corretamente”, e prever e se adaptar a ameaças futuras. A Associação Chinesa de Inteligência Artificial (caii) também estabeleceu princípios éticos. No setor privado, a Tencent enfatiza a importância de a ia ser disponível, confiável, compreensível e controlável.

Em outubro de 2019, um grupo de pesquisadores, entre eles o filósofo italiano e professor da Universidade de Oxford Luciano Floridi, publicou o artigo *The Chinese Approach to Artificial Intelligence: An Analysis of Policy, Ethics, and Regulation*, com cerca de 190 referências.¹³⁷ O foco do artigo é o contexto político-social que está moldando a estratégia chinesa de inteligência artificial, incluindo os limites de uso impostos pelos debates éticos; o conjunto amplo e diversificado de referências compõe um material de pesquisa robusto. Uma de suas conclusões é que, apesar de os princípios serem semelhantes aos do Ocidente, dada as especificidades institucionais e culturais, a ênfase da estratégia chinesa é maior nas relações de grupos e comunidades, e menor nos direitos individuais.

A China disputa com os Estados Unidos a liderança no desenvolvimento e uso das tecnologias de inteligência artificial; apesar da guerra comercial entre os dois países, os pesquisadores de ambos os países estão empenhados em garantir colaboração internacional. A primeira iniciativa conhecida nessa direção data de

2010, quando o cientista da computação Jie Tang, da Universidade de Tsinghua, Pequim, recebeu a missão de ir para os Estados Unidos e estabelecer vínculo com um renomado pesquisador norte-americano de ia. Nove anos depois, Tsinghua foi classificada como a melhor universidade de Ciência da Computação do mundo pelo *ranking* da *US News*, sendo responsável pela maior parte do 1% de artigos nos campos da matemática e da computação mais citados mundialmente entre 2013 e 2016. Para o vice-reitor da Universidade de Zhejiang em Hangzhou, We Fei, colaborador ativo no plano aidp, os desafios enfrentados pela ia não podem ser resolvidos por um país isolado: “A tarefa mais urgente é colaborar. Não podemos dizer que, para evitar a concorrência, não vamos cooperar. Em última análise, isso prejudicaria os interesses de toda a humanidade”.¹³⁸

A Universidade de Stanford há quatro anos publica o *AI Index Report* sobre o avanço anual da ia globalmente. O relatório de 2021 indicou que a China, em 2020, tornou-se líder em citações em periódicos científicos; desde 2017, a China vem ultrapassando os Estados Unidos em número de publicações científicas.¹³⁹

Em 2017, o Canadá publicou a primeira estratégia nacional de inteligência artificial; desde então, mais de 30 países publicaram documentos semelhantes. Em junho de 2020, em torno da ocde, 15 países – Austrália, Canadá, França, Alemanha, Índia, Itália, Japão, México, Nova Zelândia, República da Coreia, Cingapura, Eslovênia, Reino Unido, Estados Unidos e União Europeia – lançaram a Parceria Global em Inteligência Artificial (gpai), sob a presidência compartilhada do Canadá e da França; em dezembro de 2020, aderiram à gpai Brasil, Holanda, Polônia e Espanha. Em abril de 2021, a Comissão Europeia publicou o *Artificial Intelligence Act* (aia), primeira proposta de regulamentação do desenvolvimento e uso da ia.¹⁴⁰

Essas iniciativas associadas à inteligência artificial mundo afora formam um poderoso conjunto de referências, ponto de partida para o poder público brasileiro elaborar seus próprios caminhos de forma mais consistente e eficaz.

A autorregulamentação falhou: cabe ao poder público a tarefa de regular a ia

26.11.2021

Luciano Floridi, professor de Filosofia e Ética da Informação e diretor do *Digital Ethics Lab*, da Universidade de Oxford, é um autor “compulsivo” sobre inteligência artificial. Em artigo de 2020, Floridi alerta que um novo inverno da ia se aproxima: “O risco de todo verão de ia é que as expectativas superestimadas se transformem em uma distração em massa. O risco de todo inverno de ia é que a reação seja excessiva, a decepção muito negativa, e as soluções potencialmente valiosas sejam jogadas fora com a água das ilusões”.¹⁴¹ O alerta se justifica diante dos impasses no avanço da tecnologia e na mitigação dos riscos.

A inteligência artificial teve vários “invernos”. O primeiro, na década de 1970, período de frustração motivado, em parte, pelas previsões pessimistas dos pesquisadores Marvin Minsky e Seymour Papert sobre as redes neurais.¹⁴² Em meados da década de 1980, redescobertas geraram novo interesse na ia, mas este durou apenas até 1995. A partir de 2003, Geoffrey Hinton, Yoshua Bengio e Yann LeCun insistiram no caminho das redes neurais – com base no algoritmo *perceptron* (tipo de rede neural), proposto por Frank Rosenblatt em 1958, e no conceito de “conexionismo”, proposto pelos psicológicos David Rumelhart e James McClelland, da Universidade da Califórnia –, obtendo aprovação da academia e do mercado em 2012. Ou seja, a tecnologia está em seus primórdios.

Mesmo reconhecendo que a inteligência artificial tem potencial de contribuir no enfrentamento de problemas planetários – aquecimento global, injustiça social e migração –, Floridi argumenta que, para tal, a ia deve ser tratada “como uma tecnologia normal, nem como um milagre nem como uma praga, simplesmente como uma das muitas soluções que a engenhosidade humana conseguiu inventar”.

Cabe à sociedade investigar se as soluções de inteligência artificial vão efetivamente substituir, diversificar, complementar ou expandir as soluções anteriores, e qual o nível de aceitabilidade social dessas soluções: “Vamos realmente usar algum tipo de óculos estranho para viver em um mundo virtual ou aumentado criado por ia? Considere que hoje muitas pessoas relutam em usar óculos, mesmo quando precisam muito deles, apenas por razões estéticas”, pondera Floridi.

Como comentado em colunas anteriores, a reação inicial à disseminação da inteligência artificial, com seus potenciais danos, foi a criação de um conjunto de princípios gerais. Concebidos em 2017 na Asilomar Conference on Beneficial ai pelo Future of Life Institute, os “Asilomar Principles” são de aplicabilidade

restrita.¹⁴³ Em seguida, a autorregulamentação passou a ser propagada pelas grandes empresas de tecnologia.

Em artigo de novembro 2021, Floridi declara o fim da ilusão de autorregulamentação da indústria de tecnologia/digital.¹⁴⁴ Numa rápida retrospectiva, o filósofo relembra que até o início da década de 2000 questões éticas como privacidade, enviesamento e preconceito, moderação de conteúdo ilegal ou antiético, proteção à privacidade, *fake news* e exclusão digital eram circunscritas ao debate no âmbito acadêmico. O programa da primeira conferência da *International Association for Computing and Philosophy*, em 1986, na qual lhe coube a presidência, incluía o ensino *online* e a ideia denominada à época de “ética da computação” (posteriormente, “ética da informação” e hoje “ética digital”).

A partir de 2004, esses temas adquiriram visibilidade na opinião pública, com a consequente pressão sobre as estratégias e práticas das instituições e sobre a necessidade de criar arcabouços regulatórios. Para lidar com a crise ética, floresceu a ideia de autorregulamentação. Floridi rememora inúmeras reuniões em Bruxelas entre formuladores de políticas, legisladores, políticos, funcionários públicos e especialistas técnicos francamente favoráveis à ideia de “*soft law*”, baseada em códigos de conduta e padrões éticos da própria indústria, sem necessidade de controles externos ou imposições regulatórias. Ao longo do tempo, contudo, esse caminho não se tornou efetivo, ilustrado por experiências não exitosas.

Em junho de 2018, o Google lançou o *AI Principles*, e, em 2019, o *AI Guidebook*,¹⁴⁵ ambos com a finalidade de orientar o desenvolvimento e o uso responsáveis da inteligência artificial, sem resultados concretos. Em 2019, constituiu o Advanced Technology External Advisory Council (ateac), que reuniu oito especialistas, entre eles Luciano Floridi. Em abril de 2019, a *MIT Technology Review* publicou um artigo com um conjunto de sugestões práticas para orientar o Google nessas questões, afirmando com ironia que isso seria necessário pelo fato de o ateac ter durado apenas uma semana.¹⁴⁶

Em 2018, o Facebook criou o *Facebook Oversight Board* como um órgão independente, com o propósito de selecionar casos de conteúdo para revisão e defender ou reverter as decisões de conteúdo da plataforma; o comitê tem atualmente cerca de 20 membros, sendo o advogado e pesquisador Ronaldo Lemos o único representante da América Latina. Em janeiro de 2020, o Facebook revelou o estatuto do comitê; a série de lacunas preservava o comando da plataforma. Diante das reações contrárias, em outubro de 2020, reformulou os termos, aparentemente, atribuindo mais legitimidade ao comitê. Ainda é cedo para avaliar, mas suas funções estão longe de abarcar a dimensão dos problemas éticos da plataforma.

Diante da inoperância da autorregulamentação, emergem arcabouços regulatórios como o *Artificial Intelligence Act* (aia), da Comissão Europeia,¹⁴⁷ a regulamentação dos algoritmos de inteligência artificial pelo governo chinês, até

as iniciativas setoriais norte-americanas e o Projeto de Lei n.º 21/2020, aprovado na Câmara dos Deputados, ora em tramitação no Senado Federal.¹⁴⁸ “Chegou a hora de reconhecer que, por mais que valesse a pena tentar, a autorregulação não funcionou. A autorregulação precisa ser substituída pela lei; quanto antes melhor”, pondera Floridi. É obrigação do poder público proteger a sociedade, os cidadãos e as instituições.

Numa crítica direta aos futurologistas, apropriada a qualquer proposta que não leve em conta a complexidade do ambiente atual, Floridi argumenta que “eles gostam de uma ideia única e simples, que interpreta e muda tudo, que pode ser retratada em um livro fácil que faz o leitor se sentir inteligente, um livro para ser lido por todos hoje e ignorado amanhã. É a má dieta do *fast food* porcaria para pensamentos e a maldição do best-seller de aeroporto. Precisamos resistir à simplificação excessiva. Desta vez, vamos pensar mais profunda e extensivamente sobre o que estamos fazendo e planejando para a inteligência artificial. O exercício é denominado filosofia, não futurologia”.



IA e saúde

O ecossistema de saúde defronta-se com a disrupção de práticas tradicionais, entre outros fatores, pelas tecnologias digitais – *apps*, dispositivos móveis, ia, telemedicina, *blockchain* –, que estão transformando o acesso aos serviços de saúde, a relação médico-paciente e a relação do paciente com a própria saúde, e impactando igualmente setores periféricos como o de seguros. O reconhecimento de imagem, uma das implementações de ia mais bem-sucedidas, tem gerado resultados com alto grau de assertividade em diagnósticos dependentes de imagem, como radiologia, patologia e dermatologia. Resultados positivos estão sendo observados, igualmente, na detecção precoce de Alzheimer – ao aplicar uma das arquiteturas das redes neurais profundas, *generative adversarial network* (gan), a exames de ressonância magnética¹⁴⁹ –, na saúde mental, na personalização de medicamentos.

Apesar das evidências positivas, o setor de saúde é mais suscetível às limitações da técnica de ia (redes neurais profundas). A não explicabilidade (caixa-preta/*black-box*) do funcionamento dos modelos, por exemplo, é uma barreira à adesão dos profissionais de saúde: como recomendar um diagnóstico automatizado sem saber como o sistema chegou ao resultado? A ética é um tema sensível no setor de saúde, atento especialmente a) ao viés contido nos dados, que pode reproduzir, reforçar e ampliar os padrões de marginalização, desigualdade e discriminação existentes na sociedade; b) à designação de responsabilidade em casos de lesão/danos; e c) à privacidade de dados pessoais, nesse caso, dados sensíveis (segundo a Lei Geral de Proteção de Dados, lgpd).

Este bloco é composto de quatro artigos com foco nos benefícios da conexão médico-paciente, na *internet of bodies*, nos *chatbots* terapêuticos e nas “formas prazerosas de escape” que transcendem as redes sociais e os *games*.

Protagonismo da inteligência artificial no setor de saúde: restaurar a conexão médico-paciente

20.11.2020

A utopia pós-humanista é o projeto de melhoramento infinito das capacidades físicas, intelectuais e morais dos seres humanos, graças à convergência nbic – nanociência, biotecnologia, informática e ciências cognitivas –, gerando uma espécie humana sem as limitações de sua condição natural, como o envelhecimento biológico e a morte. Yuval Harari, em *Homo Deus*, projeta uma futura divisão do gênero humano em castas biológicas, com a constituição de uma elite privilegiada, os “super-humanos”, beneficiários da medicina voltada ao aprimoramento da condição dos saudáveis, e não à cura dos doentes.

Com a covid-19, o desafio do setor de saúde tem sido equilibrar as urgências de curto prazo com a reestruturação do setor frente às novas tecnologias – aplicativos, dispositivos móveis, telemedicina, *blockchain* –, particularmente a inteligência artificial. Esta tem o potencial de transformar o acesso aos serviços de saúde, a relação médico-paciente e a relação do paciente com a própria saúde.

O reconhecimento de imagem, técnica de ia, está sendo usado em áreas ricas em imagem, como radiologia, patologia, dermatologia, oftalmologia. Na gastroenterologia, a inteligência artificial permite identificar nos exames de colonoscopia pólipos diminutos (menores que cinco milímetros) com acurácia acima de 90% e em segundos. A fertilização *in vitro* tem tido resultados bem-sucedidos, ao selecionar embriões com mais chances de êxito (correlaciona com mais precisão doador e receptor).

A saúde mental é outra área de aplicação da inteligência artificial, como no rastreamento de depressão por meio da interação do usuário com o teclado, da voz, do reconhecimento facial, do uso de sensores e *chatbots* interativos; na previsão de medicação antidepressiva, correlacionando de forma mais assertiva medicação e paciente, e na prevenção de potencial suicídio e episódios de psicose em esquizofrênicos.

Na cirurgia robótica assistida, a inteligência artificial permite analisar dados de registros médicos pré-operatórios, guiando fisicamente o instrumento do cirurgião em tempo real durante o procedimento e reduzindo o tempo de permanência dos pacientes no hospital após a cirurgia. A ia reduziu o custo do sequenciamento genômico, importante para orientar tratamentos personalizados de câncer, e aprimorou o gerenciamento da saúde da população em geral.

Contudo, para o médico norte-americano Eric Topol, autor de *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again* (*Medicina profunda: como a inteligência artificial pode de novo humanizar a*

assistência médica), a maior contribuição da ia no setor de saúde não é reduzir erros ou volume de trabalho, nem mesmo curar doenças como o câncer, mas é a oportunidade de restaurar a conexão entre pacientes e médicos ao disponibilizar mais tempo para o atendimento.¹⁵⁰

Apesar de todo o investimento – o custo global em saúde atinge 3,5 trilhões de reais por ano, com taxa de crescimento estimada em 48% ao ano até 2023 –, o tempo alocado entre médico e paciente vem diminuindo progressivamente. Os médicos, pressionados pelos planos de saúde, pelos hospitais e clínicas, pela rotina intensa, focam cada vez mais em pedir e analisar exames e menos no paciente.

O privilégio de uma relação médico-paciente, em geral, é prerrogativa da elite; a maior parte dos cidadãos é atendida em relações precárias, particularmente na saúde pública. Nos Estados Unidos, o tempo alocado por consulta caiu de 60 minutos em 1975 para 12 minutos em 2019 (consulta de retorno, de 30 minutos para sete minutos). No Brasil, o tempo da consulta acompanha a média mundial de oito minutos, e estudos realizados na rede pública detectaram atendimentos de menos de três minutos, conflitando inclusive com o Ministério da Saúde, que define “serviço produtivo” como um atendimento de, no mínimo, 15 minutos.

Estudos apontam uma interdependência entre a qualidade da relação médico-paciente e a quantidade de erros médicos. Pesquisadores do Departamento de Medicina da Universidade do Texas sintetizaram três estudos sobre a frequência de erros de diagnóstico na população adulta norte-americana, e o resultado foi uma taxa de 5,08%, ou seja, aproximadamente 12 milhões de adultos nos Estados Unidos por ano, um em cada 20 adultos, e 50% dos erros de diagnóstico têm potencial negativo de alto risco. Estima-se que só nos Estados Unidos os falsos positivos gerem anualmente 4 bilhões de dólares de desperdício.

Apesar dos resultados efetivos, a adoção da inteligência artificial na saúde ainda é relativamente limitada. Além de restrições financeiras, de carência de profissionais capacitados e do medo de perder o emprego, a não explicabilidade (caixa-preta) do funcionamento dos modelos, legitimamente, é um fator de forte resistência.

Outro aspecto são as questões éticas: o viés contido nos dados pode reproduzir, reforçar e ampliar os padrões de marginalização, desigualdade e discriminação existentes na sociedade; nem sempre as amostras de dados usadas no treinamento dos algoritmos representam o universo das populações em estudo. Há ainda a questão da designação de responsabilidade em casos de lesão ou consequência negativa nos resultados gerados por algoritmos.

A quantidade e a qualidade dos dados é outra barreira. Em geral, os dados são coletados de pacientes internados, que já apresentam algum problema de saúde; ademais, no momento da coleta os pacientes encontram-se em condições “artificiais”, ou seja, fora de sua rotina de vida. A solução de curto prazo para treinar os algoritmos tem sido a de produzir sinteticamente grandes conjuntos de imagens em alta resolução.

O protagonismo (e poder) das *big techs* também está presente no setor de saúde. A Amazon adquiriu, em 2018, por 1 bilhão de dólares, a PillPack, serviço personalizado de entrega de medicamentos controlados em pacotes com as doses indicadas e o horário em que devem ser tomadas. Dois anos depois, anunciou a Amazon Pharmacy, com efeito imediato sobre o valor de mercado de gigantes do setor como Walgreens, segunda maior operadora de farmácias dos Estados Unidos. A Alphabet, *holding* do Google, adquiriu por 375 milhões de dólares participação societária na Oscar Health, *startup* de Nova York que oferece médicos 24 horas por dia. O Facebook, por sua vez, firmou uma parceria com a Universidade de Nova York para o desenvolvimento de máquinas de ressonância magnética com inteligência artificial.

A ia tem o potencial de reduzir custos e melhorar a eficiência geral do sistema de saúde. Desenvolver e aperfeiçoar os modelos, contudo, é uma parte menor do desafio, o principal é repensar a interação médico-paciente, requalificar os profissionais, criar novas funções, como o *chief nursing informatics officer* (cnio), acompanhar e avaliar os impactos das tecnologias sobre os pacientes, intermediando as áreas de enfermagem clínica e de tecnologia da informação.

Para proteger os direitos do paciente, é mandatório que os profissionais da saúde, os pesquisadores e os formuladores de políticas públicas conheçam os riscos e se empenhem em controlá-los.

Internet dos corpos: o corpo humano como plataforma tecnológica

19.3.2021

Marcapasso cardíaco monitorado remotamente, pâncreas artificial que monitora a glicose no sangue e fornece insulina, implantes cerebrais para tratar os sintomas de Parkinson e Alzheimer, próteses com *software* conectado aos ossos são exemplos de dispositivos ingeridos, implantados ou acoplados ao corpo humano, transformando-o em uma plataforma tecnológica. A internet dos corpos (*internet of bodies*, iob), extensão do domínio da internet das coisas (*internet of things*, iot), tem controle sobre funções vitais do corpo, convertendo-o em fonte geradora de dados pessoais com impactos na privacidade.

As pílulas inteligentes (*smart pills*) são cápsulas com sensores que percorrem o organismo em busca de sinais fora do padrão (anomalias), capazes de detectar desde doenças benignas até câncer. Criadas por Kourosh Kalantar-Zadeh, engenheiro de nanotecnologia da Universidade rmit, Austrália, as cápsulas medem pH, enzimas, temperatura, nível de açúcar e pressão arterial, oferecendo uma imagem pluridimensional do corpo humano. Nos Estados Unidos, a Agência de Alimentos e Medicamentos (Food and Drug Administration - fda) aprovou o dispositivo PillCam Colon, método de rastreamento alternativo à colonoscopia clássica invasiva; estudo dos pesquisadores israelenses Samuel Adler, do Hospital Bikur Cholim, e Yoav Metzger, da Universidade Hebraica de Jerusalém, atesta sua eficácia.¹⁵¹

A internet dos corpos contempla desde a “*wearable technology*” associada ao *fitness* (*smartwatches*, *fitness trackers*) aos microchips para fins de identificação biométrica ou concessão de autorização. A fabricante Biohax já implantou mais de 4 mil microchips para substituir chaves, milhares de suecos estão implantando microchips em seus corpos para substituir cartões-chave, identidades e até passagens de trem (de 2015 a 2018, 3 mil suecos se submeteram ao procedimento). Outro exemplo são os adesivos eletrônicos para a pele (*wearables* médicos), amplamente adotados para monitoramento cardiovascular, controle de diabetes, detecção de temperatura, suor e monitoramento de biomarcadores.

Ao permitir o monitoramento próximo e contínuo, a internet dos corpos pode dar suporte a sistemas de saúde na detecção precoce e na prevenção de doenças, ser eficaz na reabilitação de pacientes após cirurgia ou medicação, e contribuir no combate a doenças pandêmicas, como a covid-19. Os dispositivos inteligentes promovem, entre outros, o rastreamento remoto de pacientes, estilos de vida saudáveis, a medicina preventiva e de precisão, a segurança no local de trabalho. Trazem, contudo, riscos à segurança e à privacidade, suscitando novos

desafios para a governança de dados: privacidade, autonomia individual, *biohacking*, riscos de discriminação e viés, além dos problemas de modelagem e desenvolvimento.

Os alertas não são novos. Em 2018, por exemplo, Andrea M. Matwyshyn, professora de Direito e Ciência da Computação da Universidade Northeastern, publicou um artigo no *Wall Street Journal* em que reconhece os extraordinários benefícios desses dispositivos, mas alerta sobre os riscos associados, tais como a segurança física, a autonomia e o bem-estar.¹⁵²

Entre as questões éticas, talvez o maior desafio a enfrentar seja a ameaça à privacidade. Como todas as tecnologias conectadas, os dispositivos da iob geram grandes volumes de dados, particularmente dados sensíveis, categorizados pela Lei Geral de Proteção de Dados Pessoais (lgpd) como aqueles que revelam atributos pessoais, incluindo informações genéticas, biométricas e de saúde. Essa coleta, armazenamento, uso e compartilhamento em larga escala de dados pessoais, em geral, não estão previstos nas regulamentações vigentes (pelo menos em sua ampla abrangência).

O Fórum Econômico Mundial (wef), em colaboração com autoridades de saúde pública, empresas líderes de tecnologia e outras partes interessadas, está empenhado em desenvolver e testar novas abordagens para o tratamento ético do compartilhamento de dados de saúde coletados por dispositivos portáteis. Em julho de 2020, o wef publicou o relatório de autoria de Xiao Liu e Jeff Merritt, ambos os pesquisadores vinculados ao wef, *Shaping the Future of the Internet of Bodies: New Challenges of Technology Governance (Moldando o futuro da internet dos corpos: novos desafios da governança de tecnologia)*, que versa sobre as implicações desses dispositivos para a privacidade e a equidade, além do uso dos dados para fins de vigilância pública. “Estamos no início de um importante diálogo que terá grandes implicações para a saúde pública, a segurança e a economia global e também pode, em última análise, desafiar a forma como pensamos sobre nossos corpos e o que significa ser humano”, ponderam os autores.¹⁵³

O relatório examina a governança de dados da internet dos corpos nos Estados Unidos comparativamente à regulamentação da União Europeia, salientando a urgência de atualizar as abordagens de governança e as leis de proteção de dados. As regulamentações norte-americanas são setoriais, com leis específicas para distintos tipos de informação, usuário e contexto. Na Europa, como pondera o wef, o Regulamento Geral de Proteção de Dados (gdpr) é um regulamento não setorial e tecnologicamente neutro, que fornece diretrizes para os procedimentos de coleta e processamento de dados pessoais. Para os autores, tanto nos Estados Unidos quanto na União Europeia, existem lacunas entre as leis antidiscriminação e o novo risco de discriminação decorrente de inferências, perfis e agrupamentos baseados nos dados originados na internet dos corpos.

Os benefícios das novas tecnologias digitais são incontestáveis; em troca

deles, entregamos nossos dados. Desconhecemos, contudo, como, onde e para que nossos dados são usados. Desconhecemos suas implicações, presentes e futuras. A aceleração de novas descobertas e novas aplicações aumenta a lacuna entre o conhecimento dos legisladores e o dos desenvolvedores, quase que inviabilizando a atualização contínua dos arcabouços regulatórios. Desafio e impasse dos tempos atuais.

Novas fronteiras da inteligência artificial: IA emocional e *chatbot* terapêutico

28.5.2021

No livro *Klara e o Sol*, do Prêmio Nobel Kazuo Ishiguro, Klara é um “amigo artificial” (aa).¹⁵⁴ No empenho de obter autorização para abrir a “caixa-preta” e entender o funcionamento dos aas, o Sr. Capaldi argumenta com Klara: “As pessoas estão falando que vocês ficaram inteligentes demais. Elas conseguem ver o que vocês fazem. Aceitam que as suas decisões, recomendações, são sensatas e confiáveis, e quase sempre estão corretas. Mas as pessoas não gostam de não saber como vocês chegam a essas decisões”.

Gradativamente, decisões estão sendo tomadas por algoritmos de inteligência artificial, suscitando certo desconforto pelo “problema da interpretabilidade”, como os cientistas denominam a caixa-preta desses modelos (desconhecimento de como são gerados os resultados). Esse relativo desconforto, contudo, não tem impedido a crescente e diversificada aplicabilidade da ia. Uma nova fronteira é a “inteligência artificial emocional” (“ia emocional” ou “computação afetiva”); baseada em sistemas que medem, simulam e reagem às emoções humanas, a tecnologia promove uma interação “natural” e personalizada entre humanos e máquinas.

Lidando com grandes conjuntos de dados, os sistemas são capazes de captar os “dados emocionais” do usuário em tempo real, combinando dados biométricos e fisiológicos coletados por meio de análise de texto e imagem, de reconhecimento de microexpressões faciais e nuances no tom de voz, de monitoramento do movimento dos olhos e frequência cardíaca, e até do nível de imersão neurológica.

Os críticos acusam a tecnologia de ser imprecisa e preconceituosa, além de ameaçar a privacidade. Com o propósito de conscientizar sobre seus potenciais danos, uma equipe liderada por Alexa Hagerty – pesquisadora da Universidade de Cambridge, do Leverhulme Centre for the Future of Intelligence e do Ada Lovelace Institute – criou o aplicativo *emojify.info* para o usuário experimentar o reconhecimento de emoções utilizando sua própria câmera.¹⁵⁵ Financiado pela National Endowment for Science, Technology and the Arts (Nesta), o aplicativo não coleta dados pessoais, e as imagens são armazenadas no dispositivo do usuário. “É uma forma de reconhecimento facial, mas vai além, porque, em vez de apenas identificar as pessoas, afirma ler em nosso rosto as nossas emoções, os nossos sentimentos internos”, alerta Hagerty sobre a ia emocional.

A próxima geração promete ser os *chatbots* “assistentes terapêuticos” ou “agentes de conversação automatizada”. Entre os projetos em curso, destaca-se o

chatbot Woebot, inicialmente disponível apenas no Facebook, com interação via Messenger, e agora podendo ser baixado gratuitamente na Apple Store e no Google Play (Woebot – Your Self-Care Expert).

Desenvolvido por pesquisadores da Universidade de Stanford com a missão de “tornar a saúde mental radicalmente acessível”, o Woebot oferece terapia virtual com uso de inteligência artificial. O foco do aplicativo são as terapias cognitivo-comportamentais (tcc). Por 10 minutos diários de atendimento a 39 dólares ao mês, os “pacientes” são estimulados a contar suas reações emocionais aos eventos do cotidiano e, a partir delas, identificar as armadilhas psicológicas que originam estresse, ansiedade e depressão. Segundo Michael Evers, ceo da Woebot Health Inc.: “Estamos construindo instrumentos clínicos e terapêuticas digitais que automatizam o conteúdo e o processo de terapia, adaptam-se aos sintomas e à gravidade e fornecem a intervenção certa para a pessoa certa no momento certo. Eles estão na vanguarda de uma nova geração de intervenções que realmente tornarão a saúde mental radicalmente acessível a todos”.¹⁵⁶

Numa pesquisa publicada na revista *JMIR Formative Research*, a Woebot Health pondera que os resultados são “comparáveis aos serviços tradicionais prestados por humanos em todas as modalidades de tratamento e foram significativamente mais altos do que programas computadorizados de terapia cognitivo-comportamental (ccbt)”.¹⁵⁷ Com base numa amostra de 36.070 usuários, entre 18 e 78 anos, 57,48% do sexo feminino, o estudo conclui que: “Embora os vínculos sejam frequentemente presumidos como domínio exclusivo das relações terapêuticas humanas, nossos achados desafiam a noção de que a terapêutica digital é incapaz de estabelecer um vínculo terapêutico com os usuários”. Esse vínculo, segundo apurado pela pesquisa, foi estabelecido num intervalo de apenas três a cinco dias, mantendo-se estável ao longo do tempo.

A fundadora e presidente da Woebot Health, Alison Darcy, psicóloga e cientista da computação, em suas pesquisas na Universidade de Stanford, contou com a colaboração do reconhecido especialista em inteligência artificial Andrew Ng, ambos interessados em como a ia poderia ajudar as pessoas com problemas de saúde mental. Para Darcy: “A capacidade de estabelecer um vínculo, e fazê-lo com milhões de pessoas ao mesmo tempo, é o segredo para desvendar o potencial da terapêutica digital como nunca antes. Estamos entusiasmados por estar na vanguarda desta linha de pesquisa”.¹⁵⁸

A covid-19, com seus efeitos perversos sobre a saúde mental, acelera essas tendências. Para dar conta do aumento da demanda por serviços de psicólogos, psicanalistas e psiquiatrias, e igualmente pelas restrições impostas pela quarentena, proliferam plataformas habilitadas por tecnologias baseadas em vídeo e mediadas pela internet – programas habilitados de realidade virtual e ia. A *Frontiers in Psychiatry*, revista científica revisada por pares, publicou um estudo sobre o uso de terapia cognitivo-comportamental via internet; com 2.866 participantes em três categorias – transtornos de ansiedade, depressão e outros –,

observou resultados positivos, particularmente entre os nativos digitais (crianças, adolescentes e adultos jovens): 65,6% de todos os pacientes responderam, e cerca de um terço alcançou a remissão.¹⁵⁹

A técnica que permeia atualmente a maior parte das aplicações de inteligência artificial (redes neurais profundas/*deep learning*) ainda está em seus primórdios, concretizada efetivamente na última década. Há certo encantamento por parte da sociedade sobre seus benefícios, que, efetivamente, são extraordinários, mas é imprescindível a consciência de suas limitações e seus riscos, particularmente quando se associa ia e saúde mental.

Nem tudo é tecnologia: são múltiplas as formas prazerosas de escape

21.1.2022

Ficção: Rue, personagem principal da série *Euphoria*, interpretada pela atriz, cantora e compositora Zendaya, foi diagnosticada ainda bem jovem com transtorno obsessivo compulsivo (toc), transtorno de déficit de atenção (tdah) e transtorno de ansiedade geral (tag), sendo medicada por anos com drogas lícitas. Aos 17, Rue é dependente de drogas lícitas e ilícitas. A série da hbo, criada e produzida por Sam Levinson a partir de sua própria experiência de adolescente e baseada na série original israelense de 2012, retrata o cotidiano de um grupo de adolescentes do ensino médio norte-americano e os excessos de consumo de drogas, álcool, sexo e mídias sociais. Elogiada pela crítica, a série tem 80% de aprovação no Rotten Tomatoes (site norte-americano agregador de críticas de cinema e televisão, fundado em 1998 e desde 2011 de propriedade da Warner Bros.).

Realidade: David, norte-americano de 20 anos, no segundo ano da faculdade, recorreu ao serviço de saúde mental da universidade, buscando ajuda para ansiedade e baixo desempenho escolar. Após uma consulta e um teste de menos de cinco minutos, David foi diagnosticado com transtorno de déficit de atenção (tdah) e transtorno de ansiedade generalizada (tag). O psicólogo que administrou o teste recomendou um psiquiatra, que receitou Paxil para tratar depressão e ansiedade, e Adderall para tratar tdah. Após anos de medicação, David tornou-se dependente de drogas lícitas e ilícitas.

David (nome fictício) é paciente na clínica para dependentes da Dra. Anna Lembke (que se reconhece como dependente de leitura de romances), médica diretora da Stanford Addiction Medicine e chefe da Stanford Addiction Medicine Dual Diagnosis Clinic, premiada por pesquisas em doenças mentais e membro de conselhos de organizações estaduais e nacionais. No livro *Nação dopamina: por que o excesso de prazer está nos deixando infelizes e o que podemos fazer para mudar*,¹⁶⁰ Lembke pondera: “somos atraídos para qualquer uma das formas agradáveis e fuga, que agora estão disponíveis para nós: coquetéis da moda, a câmara de ressonância da mídia social, maratona de *reality shows*, uma noite de pornô pela internet, batatas fritas e *fast food*, video games envolventes, romances baratos de vampiro... A lista realmente não acaba”.

Segundo Lembke, atualmente, mais de um em cada 10 norte-americanos toma antidepressivo (110 pessoas por 1.000) – Islândia (106/1.000), Austrália (89/1.000), Canadá (86/1.000), Dinamarca (85/1.000), Suécia (79/1.000) e Portugal (78/1.000). Entre 25 países, a menor proporção é da Coreia (13/1.000).

Em quatro anos, o consumo de antidepressivos aumentou 46% na Alemanha e 20% na Espanha e em Portugal; as prescrições de estimulantes (Adderall, Ritalina) dobraram nos Estados Unidos entre 2006 e 2016, inclusive em crianças menores de 5 anos; e entre 1996 e 2013, o número de adultos que receberam uma receita de benzodiazepínicos aumentou 67%. Em paralelo, entre 2002 e 2013, a dependência de álcool aumentou 50% em adultos com mais de 65 anos e 84% em mulheres.

Relatório da Association of Schools and Programs of Public Health (aspph),¹⁶¹ de novembro de 2019, concluiu: a tremenda expansão de prescrição de opioides lícitos aumentou a dependência e a transição para opioides ilícitos, e em seguida, exponencialmente, o número de overdoses. Overdoses de opioides mataram mais norte-americanos do que armas ou acidentes de carro.

Segundo o Departamento de Saúde e Serviços Humanos dos Estados Unidos, países com piores perspectivas econômicas são mais propensos a ter taxas mais altas de prescrições de opioides, hospitalizações relacionadas a opioides e mortes por overdose de drogas. Segundo a mesma instituição, norte-americanos que usam o Medicaid, seguro de saúde financiado pelo governo federal para pobres e vulneráveis, recebem duas vezes mais analgésicos opioides do que pacientes não Medicaid e morrem de opioides de três a seis vezes mais que pacientes não Medicaid.¹⁶²

Estudo dos economistas de Princeton Anne Case e Angus Deaton constatou que os norte-americanos brancos de meia-idade sem diploma universitário estão morrendo mais jovens que seus pais, avós e bisavós; as três principais causas de morte nesse grupo são overdose de drogas, doença hepática relacionada ao álcool e suicídio, mortes denominadas por Case e Deaton de “mortes por desespero”.¹⁶³

No Brasil existem diversos estudos sobre dependência de drogas lícitas e ilícitas. Um dos mais recentes, realizado na Universidade de São Paulo, sobre pacientes internados por traumas em serviços de emergência na cidade – “Use of alcohol and illicit drugs by trauma patients in Sao Paulo, Brazil” –, constatou a prevalência de substâncias psicoativas em 31,4% dos pacientes, em especial álcool e cocaína, o que provavelmente contribuiu para um terço dos acidentes de carro com lesões.¹⁶⁴

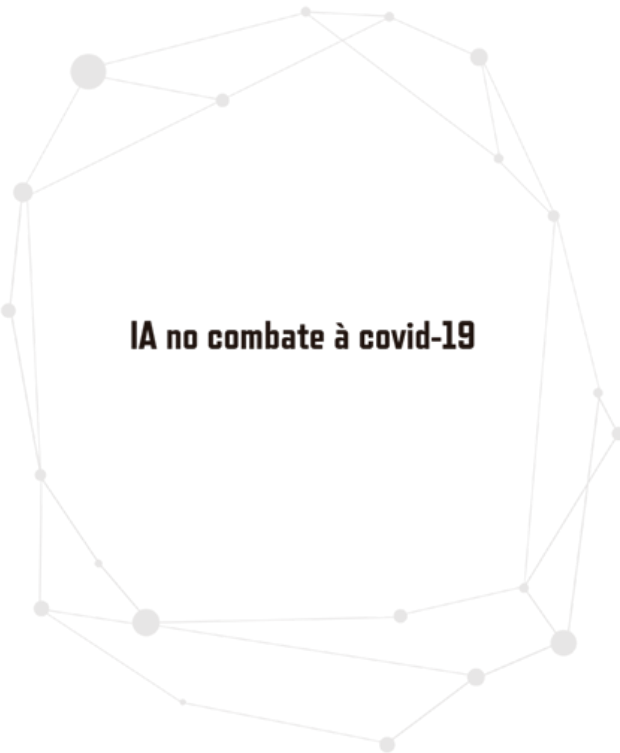
Esse sucinto panorama do nível de dependência mundial das mais variadas “formas prazerosas de escape” serve como argumento da necessidade de focar nas causas dos problemas sociais, e não nos seus efeitos.¹⁶⁵ Sim, os modelos de negócio das redes sociais e dos *games online* são concebidos, design e dinâmica, para reter os usuários nas respectivas plataformas, ou seja, “viciá-los”. Contudo, a título de especulação – ou investigação dos especialistas –, a causalidade não poderia ser inversa? Parte do sucesso dessas plataformas não poderia decorrer da propensão dos usuários a buscar “formas prazerosas de escape”?

No livro *Games viciam: fato ou ficção?*, os autores, entre eles as psicólogas da Pontifícia Universidade Católica de São Paulo (puc-sp) Ivelise Fortim e Maria

Thereza Alencar Lima, com base em evidências científicas, dedicam-se a responder à pergunta-título.¹⁶⁶ Segundo os autores, estudos identificaram prevalência de dependência de *games* em entre 0,3 e 3% da população, concentrados em países asiáticos, especialmente Japão, China e Coreia do Sul; nos países ocidentais o percentual é bem menor (não existe pesquisa no Brasil). Parcela muito significativa dos usuários com dependência em *games* apresentam concomitantemente outros problemas emocionais. “Quando relatado na mídia, o foco muitas vezes recai sobre os raríssimos usuários de jogos *online* que morreram em suas cadeiras por falta de comida, água, ou por permanecerem inúmeras horas na mesma posição, provocando trombose”, ponderam os autores. (Gerou polêmica entre os especialistas a inclusão, em 2018, pela Organização Mundial de Saúde [oms] – 11ª edição da Classificação Internacional de Doenças [cid-11] – da “gaming disorder” na categoria de dependências comportamentais.)

Lembke lista as principais causas da crescente dependência: fácil acesso (*online* e *offline*), tempo livre e falta de opções de lazer, mudanças culturais e comportamentais, mudança na dinâmica familiar. Um dia típico de um trabalhador médio nos Estados Unidos pouco antes da Guerra Civil (1861-1865), por exemplo, consumia de 10 a 12 horas/dia; atualmente, o número típico de horas trabalhadas nos Estados Unidos é de 3,8 horas/dia (número semelhante ao de países de alta renda). O tempo livre é inversamente proporcional ao status socioeconômico (os adultos nos Estados Unidos sem diploma de ensino médio têm 42% mais tempo livre do que os adultos com diploma de bacharelado ou mais) e, conseqüentemente, ao acesso a opções de lazer.

A covid-19 não é a única epidemia da sociedade atual, como também não o são as redes sociais e os *games online*, cenário agravado pela crescente desigualdade – desde o início da pandemia um novo bilionário surgiu a cada 26 horas, a riqueza dos 10 homens mais ricos do mundo dobrou, a situação de 99% da humanidade piorou e mais de 160 milhões foram empurrados para a pobreza.¹⁶⁷ Regularizar os modelos de negócio baseados em dados é crítico, sem negligenciar, como diz Evgeny Morozov, que “o verdadeiro inimigo não é a tecnologia, mas o atual regime político e econômico”.¹⁶⁸



IA no combate à covid-19

A canadense BlueDot, plataforma de inteligência artificial (ia) que rastreia doenças infecciosas, sinalizou, nove dias antes da Organização Mundial da Saúde (oms), casos de pneumonia incomum detectados no mercado em Wuhan, cidade chinesa de origem da covid-19. Pesquisadores mundo afora desenvolveram soluções de ia no enfrentamento da pandemia para rastrear o surto, diagnosticar pacientes, desinfetar áreas e acelerar o desenvolvimento de vacinas. Em fevereiro de 2020, a Missão oms-China divulgou um relatório sobre a resposta inicial à pandemia, ressaltando o protagonismo da ia.¹⁶⁹

Ao longo do tempo, contudo, a prática evidenciou as limitações inerentes às técnicas de ia. Baseadas em extrair padrões de grandes conjuntos de dados, a qualidade dos dados é fator crítico.¹⁷⁰ Como pondera a oecd: “Embora a covid-19 esteja mostrando que a ia pode oferecer benefícios em situações de crise, também atesta o *hype* em torno da ia”.¹⁷¹

Pela dimensão da crise e a percepção inicial favorável a essas tecnologias, a coluna dedicou seis artigos à covid-19 abordando as ameaças à privacidade e à liberdade (sem deixar de reconhecer os benefícios), as restrições ao uso de dados sensíveis e a vida mediada pelas máquinas durante a quarentena, e introduzem o conceito de “epidemiologia digital” com os procedimentos inovadores de monitoramento de epidemias. O sexto e último artigo, diante das evidências, pondera sobre as limitações dos modelos de ia que comprometem os resultados.

Sistemas de vigilância digital: combatem o coronavírus, mas ameaçam a privacidade e a liberdade

27.3.2020

O filósofo sul-coreano Byung-Chul Han, professor na Universidade de Berlim, é atualmente o pensador asiático de maior sucesso no Ocidente. Crítico das redes sociais e da sociedade contemporânea, suas declarações “viralizam”, e suas publicações figuram entre os conteúdos mais lidos. Em 22 de março, o filósofo publicou no *El País* um artigo intitulado “O coronavírus de hoje e o mundo de amanhã”, com enorme repercussão nas mídias digitais, inclusive no Brasil.¹⁷²

Byung-Chul Han, aparentemente, não é crítico em relação aos sistemas de vigilância e controle. O filósofo argumenta que uma das vantagens da Ásia no combate à epidemia são seus sistemas de vigilância digital, particularmente na China e na Coreia do Sul. Segundo ele, o sistema da China permite lidar melhor com a epidemia, por outro lado, as leis de proteção de dados agravam o quadro da Europa: “Pode-se dizer que as epidemias na Ásia não são combatidas apenas por virologistas e epidemiologistas, mas sobretudo por cientistas da computação e especialistas em *big data*. Uma mudança de paradigma sobre a qual a Europa ainda não aprendeu”.

Com seus 200 milhões de câmaras equipadas com sofisticadas tecnologias de reconhecimento facial, na China o controle é praticamente total. Sua eficiência, em parte, é em função do compartilhamento irrestrito dos dados dos cidadãos entre os provedores de telefonia e internet e os órgãos do governo. No auge da crise do coronavírus, o mundo assistiu, alguns perplexos, agentes do governo com câmeras equipadas com medidores de temperatura do corpo humano interceptando chineses na entrada do metrô, por exemplo, e os conduzindo para triagem. Os drones, igualmente espalhados pelas ruas, detectavam se um cidadão, supostamente em quarentena, saiu de casa; nesse caso, os próprios drones ordenavam seu retorno imediato.

Um novo *software*, obrigatório nos *smartphones* chineses, identifica quem deve ou não ser colocado em quarentena ou liberado para o transporte público. A primeira experiência do programa aconteceu na cidade de Hangzhou (sede da Alibaba e de seu “braço” financeiro, Ant Financial), hoje já implantado em 200 cidades e, em breve, em todo o país. Criado pela gigante de tecnologia Alibaba, o *software* recebeu a denominação de “Código de Saúde Alipay”: a cor do código indica o estado de saúde do dono do dispositivo. Após o usuário preencher um formulário na Alipay com detalhes pessoais, o *software* gera um código qr em uma das três cores: verde, liberado; amarelo, em casa por sete dias; e vermelho,

quarentena de duas semanas.

Em Hangzhou é quase impossível se locomover sem mostrar o código Alipay (98,2% da população da província de Zhejiang, cuja capital é Hangzhou, deram código verde; quase 1 milhão de pessoas, códigos amarelos e vermelhos). O jornal *The New York Times* analisou o código do *software* e descobriu, contudo, que o sistema também compartilha informações com a polícia, parceiro-chave no projeto, representando uma nova modalidade de controle social (ainda não está claro se é conjuntural ou permanente).¹⁷³

Como usual na China, não há transparência sobre os critérios de classificação do *software*. Visando facilitar o entendimento dos ocidentais, os autores do artigo do *The New York Times* usaram uma analogia: “Nos Estados Unidos, seria semelhante ao Centro de Controle e Prevenção de Doenças (cdc) usar os aplicativos da Amazon e do Facebook para rastrear o coronavírus e, em seguida, compartilhar silenciosamente as informações do usuário com o escritório do xerife local”. Para Maya Wang, pesquisadora da Human Rights Watch na China o surto de coronavírus está provando ser um desses marcos na história da disseminação da vigilância em massa na China: “Uma vez que esses sistemas estão em vigor, os envolvidos em seus desenvolvimentos defendem sua expansão ou seu uso mais amplo, um fenômeno conhecido como ‘*mission creep*’”.¹⁷⁴ No entanto, corroborando a visão de Byung-Chul Han, parece que os chineses não estão preocupados com os efeitos sobre a privacidade num cenário pós-epidemia: as empresas e o governo já possuem todos os seus dados, e o governo já os utiliza para exercer controle sobre a população.

A efetividade dos sistemas de vigilância sobre a epidemia da covid-19 tem estimulado governos de vários países a flexibilizar as legislações de proteção de dados pessoais (ao menos enquanto durar a epidemia). Em Israel, num primeiro teste, cerca de 400 pessoas receberam uma notificação do Ministério da Saúde com o alerta de proximidade de alguém com resultado positivo para o vírus. Nos Estados Unidos, segundo o *The Washington Post*, o governo solicitou ao Facebook e ao Google acesso aos dados de localização de seus usuários, visando avaliar se eles estão respeitando as distâncias recomendadas; e o Departamento de Saúde e Serviços Humanos (Department of Health and Human Services) anunciou a suspensão de multas por violações relacionadas aos padrões de privacidade de dados de saúde, facilitando a atuação dos médicos. O governo inglês firmou uma parceria com a operadora de telefonia móvel O2: por meio da análise de dados anônimos de localização de *smartphone*, pretende checar se seus usuários estão seguindo as diretrizes de combate ao vírus.

A Assembleia Global de Privacidade (gpa) – principal fórum global de proteção de dados e autoridades de privacidade, fundada em 1979 e com mais de 130 membros – identificou mudanças relacionadas à privacidade de dados em pelo menos 27 países. A gpa está atenta para conter potenciais abusos. Ao tomarem essas decisões, as autoridades de proteção de dados em todo o mundo

estão priorizando vidas em detrimento da privacidade.

Entre inúmeros problemas que a sociedade enfrentará no pós-epidemia, inclui-se o desafio de retroceder os sistemas de vigilância, ao menos, aos níveis anteriores à epidemia, evitando ampliar o controle dos governos sobre seus cidadãos, clara ameaça à privacidade e à liberdade.

O protagonismo da inteligência artificial no combate à covid-19

10.4.2020

Um grafite em Hong Kong vaticina: “Não podemos voltar ao normal, porque o que era normal era exatamente o problema”. A covid-19 escancarou as deficiências do mundo, a desigualdade se tornou assustadoramente visível, colocando a sociedade diante de vários dilemas. Ainda é cedo para prever como será o mundo pós-coronavírus, qual será o grau e a extensão dos impactos na economia, na sociedade e na vida dos indivíduos. Provavelmente, não transitaremos pelos espaços públicos com a mesma desenvoltura anterior à epidemia (pelo menos não tão cedo).

A ciência está em alta, principalmente a ciência apoiada nas tecnologias digitais. Nesse cenário, destaca-se o protagonismo da inteligência artificial; seus algoritmos são treinados para entender as “regras”, tornando-se aptos a realizar previsões com maior velocidade e acurácia. A eficiência dos modelos depende da quantidade e da qualidade dos dados de treinamento; no combate à covid-19, proliferam iniciativas para compor e disponibilizar grandes bases de dados. O propósito é identificar padrões não visíveis, indícios, entre outros eventos, da probabilidade de contaminação.

A China e a Coreia do Sul utilizaram tecnologias de rastreamento para identificar pacientes contaminados, adotadas posteriormente em vários países. Nos Estados Unidos, pesquisadores do mit desenvolveram uma tecnologia de rastreamento de infectados pela covid-19 por meio do celular. O aplicativo Private Kit mapeia os contatos pessoais entre os usuários (“rastreamento de contatos”), visando antecipar potenciais áreas de contaminação. Para preservar a privacidade, os dados coletados são anonimizados, impossibilitando a individualização (o rastreamento é realizado sem o conhecimento ou a permissão explícita do usuário). A equipe do mit está em negociação com a oms e com o Departamento de Saúde e Serviços Humanos dos Estados Unidos; o desafio é conquistar ampla adesão, acumulando uma extensa base de dados (um dos caminhos é através de alianças; a primeira foi firmada com a Clínica Mayo). O projeto do mit ganhou a adesão de engenheiros do Facebook (doação de tempo gratuito). Em paralelo, o Facebook e o Google estão cooperando com uma tarefa da Casa Branca na análise de outras tecnologias de rastreamento e localização para mitigar a epidemia.

Os sistemas inteligentes também são úteis em processos de triagem para estimar o número de leitos necessários nos hospitais, avaliar a prioridade de pacientes em unidade de terapia intensiva (uti) e determinar a priorização de

pacientes na UTI para o uso de aparelhos de ventilação pulmonar. Com base nos dados, os modelos de IA dão suporte a decisões médicas, agregando velocidade e precisão, dois elementos valiosos numa crise dessa dimensão. Outra aplicabilidade da inteligência artificial relacionada à COVID-19 são os robôs autônomos, treinados para executar tarefas específicas nos centros de saúde, tais como desinfetar áreas e entregar materiais e medicamentos.

No Brasil, como no resto do mundo, instituições e profissionais de saúde estão colaborando no enfrentamento da epidemia. Um consórcio, por exemplo, formado por hospitais, grupos de saúde e empresas de tecnologia, com a participação do Hospital das Clínicas de São Paulo, está testando um algoritmo de inteligência artificial para diagnóstico com base na análise de imagens de tomografia de pulmão; um grupo de 450 radiologistas brasileiros, de instituições públicas e privadas, organizou-se para compartilhar informações sobre IA.

Pesquisadores do Hospital Renmin, da Universidade de Wuhan, da Wuhan EndoAngel Medical Technology Company e da Universidade de Geociências da China anunciaram um modelo de IA para detectar o vírus com 95% de precisão.¹⁷⁵ A *startup* canadense DarwinAI e pesquisadores da Universidade de Waterloo desenvolveram o Covid-Net de código aberto, uma rede neural convolucional criada para detectar a COVID-19 em imagens de raios X.¹⁷⁶

O Kaggle – comunidade *online* de cientistas de dados, fundada em 2010 e adquirida pelo Google em 2017 – propôs um desafio em torno de 10 questões-chave relacionadas ao coronavírus, incluindo perguntas sobre fatores de risco e tratamentos que não envolvem medicamentos, propriedades genéticas do vírus e esforços para desenvolver vacinas. Participam do projeto a Chan Zuckerberg Initiative (instituição fundada por Mark Zuckerberg e Priscilla Chan com foco em projetos sociais) e o Centro de Segurança e Tecnologias Emergentes da Universidade de Georgetown.¹⁷⁷

A comunidade científica mundo afora tem produzido uma quantidade inédita de artigos; desde dezembro, foram publicados mais de 2 mil artigos sobre a COVID-19. O processamento de linguagem natural (*natural language processing*, NLP), subárea da inteligência artificial voltada à compreensão automática da linguagem humana, permite ler artigos científicos e extrair informações úteis em quantidade e velocidade inalcançáveis pelo ser humano, mantendo atualizados os profissionais da saúde e, consequentemente, acelerando a descoberta de tratamentos e fatores de riscos desconhecidos. A canadense CiteNet, em face da COVID-19, abriu seu sistema de consulta ao público, disponibilizando sua base de dados, continuamente atualizada, de pesquisas e literatura científica.

O enfrentamento de epidemias por meio de análise de dados não é novidade; a primeira experiência em larga escala ocorreu em fevereiro de 2008, quando o Centro de Controle e Prevenção de Doenças (CDC) identificou um crescimento de casos de gripe no leste dos Estados Unidos; na ocasião, o Google declarou ter detectado um aumento nas consultas sobre os sintomas da gripe duas

semanas antes do lançamento do relatório. A partir dessa experiência, sua unidade filantrópica criou um sistema de alerta, o Google Flu Trends. A metodologia estabeleceu correlações entre a frequência de digitalização de certos termos de busca no Google e a disseminação da gripe ao longo do tempo e do espaço, identificando regiões específicas em tempo real (posteriormente, especialistas questionaram a veracidade/eficácia desse sistema).

No caso da epidemia da covid-19, o primeiro alerta veio da canadense BlueDot, em 31 de dezembro de 2019, quando detectou uma nova forma de doença respiratória na região do mercado de Wuhan, na China (nove dias antes do primeiro comunicado da oms). Por meio de análise avançada de dados com tecnologias de inteligência artificial (“varrendo” milhares de fontes: documentos de autoridades, publicações médicas, relatórios climáticos, cerca de 100 mil jornais por dia em 65 idiomas, postagens em redes sociais), seu *software* de risco identifica potenciais ameaças de doenças infecciosas, capacitando governos, profissionais e empresas da área de saúde; suas soluções rastreiam, contextualizam e antecipam riscos de epidemias.

A missão da oms na China, num relatório de 40 páginas divulgado em março, documentou a relevância do *big data* e da inteligência artificial na estratégia chinesa, particularmente no rastreamento de contatos para monitorar a propagação da contaminação e no gerenciamento de populações prioritárias (idosos e chineses com doenças prévias de risco).

A covid-19 transformou nosso cotidiano num filme de ficção científica; a devastação social não permite ver “o lado bom” da epidemia, mas é certo que teremos avanços significativos no campo da inteligência artificial, com novas descobertas e novas aplicações.

As limitações da inteligência artificial no enfrentamento da covid-19

24.4.2020

No livro *Os sertões* (1902), marco da literatura brasileira, o escritor e jornalista Euclides da Cunha narra a sangrenta Guerra de Canudos no interior da Bahia, liderada por Antônio Conselheiro. O livro expõe de forma realista a vida dos jagunços e dos sertanejos. Descrevendo o cotidiano no acampamento de Canudos, o escritor cria uma expressão muito apropriada aos tempos atuais: “A vida normalizara-se naquela anormalidade”. A covid-19 implicou violações inimagináveis às liberdades individuais: o direito de circular pelas ruas das cidades, agora regulado pelo poder público. Provavelmente, até o final do ano teremos “quarentenas intermitentes” (libera/flexibiliza, aumenta a contaminação, retrocede) e um cenário econômico volátil. Nessa vida suspensa, parte das expectativas está depositada nas tecnologias de inteligência artificial.

Enquete realizada pelo Aspen Institute, da Alemanha, em recente *live*, indicou os campos com maior potencial para a ia no combate à epidemia: 36% em rastreamento (aplicativo de monitoramento de movimentação), 22% no desenvolvimento de vacinas e 20% na previsão de propagação do vírus. No campo da ética, 42% dos desafios estão no gerenciamento de recursos hospitalares e na triagem de pacientes e, empatados com 21%, o reconhecimento de imagem e os aplicativos de rastreamento.

Pressionados pela urgência, pesquisadores mundo afora estão empenhados em criar modelos para detectar a contaminação, prever a evolução de pacientes contaminados, identificar grupos de risco e descobrir vacinas e medicamentos. Existem, atualmente, 115 estudos de vacina contra a covid-19 acontecendo no mundo, sendo 78 registrados e cinco já em testes clínicos; nessa semana, a BioNTech e a Pfizer aprovaram junto à autoridade reguladora alemã – o Instituto Paul-Ehrlich – o ensaio clínico de fase 1/2 do programa compartilhado de vacinas bnt162 da BioNTech; essa fase inclui testes em cerca de 200 indivíduos saudáveis, com idades entre 18 e 55 anos, com o objetivo de determinar a dose ideal, além de avaliar a segurança e a imunogenicidade da vacina. No entanto, segundo o biólogo brasileiro Atila Iamarino, no melhor cenário as vacinas estarão disponíveis em meados de 2021 (tempo recorde comparado com o desenvolvimento da vacina do ebola, por exemplo, que demorou cinco anos).

Parte dos estudos publicados, contudo, não observou as regras científicas, ou seja, não foi revisado por outros especialistas (“revisão por pares”). Em 31 de março, um grupo de pesquisadores europeus publicou uma análise crítica na conceituada revista científica *British Medical Journal*, intitulada “Prediction

Models for Diagnosis and Prognosis of Covid-19 Infection: Systematic Review and Critical Appraisal” (“Modelos de previsão para diagnóstico e prognóstico de infecção por Covid-19: revisão sistemática e avaliação crítica”),¹⁷⁸ identificando inconsistências nos resultados de 31 modelos utilizados em 27 estudos: “Todos os modelos relataram desempenho preditivo bom a excelente, mas todos foram avaliados como tendo alto risco de viés devido a uma combinação de relatórios inadequados e conduta metodológica deficiente para seleção dos participantes, descrição do preditor e métodos estatísticos utilizados”, ponderam os pesquisadores europeus.

O alto risco de viés identificado questiona a confiabilidade de suas previsões: as amostras de pacientes utilizadas não representavam a população-alvo dos modelos, ou eram relativamente pequenas (poucos dados em modelos de *machine learning* podem funcionar em previsões mais simples, como prever a batida de falta sem barreira num jogo de futebol, em que parte das variáveis é conhecida); ademais, apenas um estudo utilizou dados coletados de pacientes fora da China, o que impacta negativamente a acurácia do modelo em razão, entre outros fatores, das diferenças entre os sistemas de saúde dos países, e o biótipo e o comportamento da população.

Segundo a análise: “Como esperado, nesses primeiros estudos com modelos de previsão relacionados à covid-19, os dados clínicos de pacientes contaminados ainda são escassos e limitados a dados da China, Itália e registros internacionais. Com poucas exceções, o tamanho da amostra disponível e o número de eventos para os resultados de interesse foram limitados”. Os estudos também não foram transparentes com relação às etapas dos testes. Essas e outras falhas prenunciam que as previsões são, provavelmente, menos assertivas do que os relatos.

Os modelos estatísticos de previsão têm uma limitação intrínseca que é o fato de prever o futuro com base em dados do passado (inferências de tendências passadas). Essa limitação não ocorre quando se trata de modelos preditivos com séries não temporais – *computer vision*; processamento de linguagem natural (pnl), técnica de ia que ajuda as máquinas a analisarem, interpretar e gerarem textos; e outros sistemas fechados –, mas é uma questão real em modelos de série temporais em tempo real (caso da epidemia da covid-19). A falta de acuidade das previsões impacta negativamente a abordagem do combate à epidemia e o tratamento dos infectados (procedimentos, remédios e vacinas).

A covid-19 confirmou que a inteligência artificial é a tecnologia mais promissora do século xxi, mas, simultaneamente, alertou para suas limitações. Como todos os modelos estatísticos de probabilidade, seus resultados indicam apenas a chance de algo acontecer (e quando); no caso dos modelos de aprendizado de máquina e redes neurais (*deep learning*), a variável “dado” – quantidade, diversidade e qualidade – é um componente sensível do modelo, e onde reside a principal causa das falhas detectadas nos modelos analisados. Isso se dá, entre outros fatores, em função do crescimento exponencial da contaminação;

do período de contaminação relativamente curto; da trajetória da contaminação variar por país e por região, de acordo com as medidas adotadas, o clima, o comportamento da população; e dos métodos de contabilização de pacientes positivos e mortes não serem universais.

O coronavírus está sendo chamado de “acelerador de futuros”. Visivelmente, a covid-19 está acelerando o avanço da inteligência artificial, incluindo o melhor entendimento de suas ainda muitas limitações.

O cenário futurista de E. M. Forster e a covid-19: a vida cotidiana transcorre mediada pela Máquina

8.5.2020

O escritor britânico E. M. Forster publicou, em 1909, a novela *A máquina parou*¹⁷⁹, que retrata um cenário futurista distópico, em que a vida humana é mediada por uma máquina. A Máquina evita qualquer deslocamento: caiu um objeto no chão? O chão “se ergue” e traz o objeto até você. A professora Vashti leciona por meio da Máquina, interage com o filho Kuno pela Máquina, alimenta-se pela Máquina, consome pela Máquina. Um dia, Kuno, em conflito com o *status quo*, pede à mãe: “Quero que você venha me ver”, ao que Vashti responde: “Mas eu posso vê-lo! O que mais você quer?”. “Não quero vê-la através da Máquina”, implora Kuno. Soa familiar?

Os personagens de Forster vivem numa pequena sala de formato hexagonal, como a célula de uma abelha. “A iluminação não vem de alguma janela, nem de uma lâmpada, no entanto um brilho suave espalha-se por todo o lugar. Não existem aberturas para ventilação e no entanto o ar é fresco.” A covid-19, como na novela de Forster, encurtou o espaço físico; passamos o dia em frente a uma máquina, por ela nos informamos, socializamos, trabalhamos, consumimos e nos divertimos. O coronavírus redefiniu, igualmente, o espaço urbano, unificando três “cidades”: a “cidade-moradia”, a “cidade-trabalho” e a “cidade-consumo”. Nossa fisicalidade mudou de natureza!

Conectados através de uma placa redonda iluminada, Vashti comenta com o filho Kuno: “Não me sinto bem”. “De imediato um enorme aparato desceu do teto sobre ela, um termômetro foi automaticamente colocado sobre seu coração. Ficou deitada, inerte. Compressas frias envolviam sua testa. Kuno havia telegrafado para o médico dela.” Não tão de imediato nem tão eficientes, cresce o número de empresas que oferecem aos seus funcionários, durante a quarentena, os serviços de telemedicina (consultas médicas *online*), como benefício extra ao plano de saúde.

Na novela de Forster, a Máquina exerce controle absoluto sobre os habitantes desse planeta imaginário, a vida é regida pelo Livro da Máquina, publicação do Comitê Central, no qual estão contidas as instruções sobre todas as contingências, incluindo botões a serem acionados em cada necessidade ou desejo. No planeta Terra, a vida extraordinária da covid-19 acelerou a implantação de mecanismos de controle cada vez mais sofisticados, que não sabemos se serão temporários ou definitivos. Com funcionalidades distintas, nem todos ameaçam a privacidade de seus usuários, mas todos têm em comum a opacidade, pelo menos para o grande público. Vejamos três modelos de

monitoramento da epidemia em curso.

A Apple e o Google formaram uma parceria inédita com potencial de monitorar cerca de um terço da população mundial. A tecnologia de “rastreamento de contato” será incorporada aos sistemas operacionais ios e Android. Usando sinais *bluetooth*, o aplicativo indica quão perto você chegou de pacientes diagnosticados com covid-19 (a eficiência do sistema depende de os usuários positivos inserirem os dados no aplicativo). A iniciativa é controversa, porque envolve o compartilhamento de informações confidenciais de saúde, além de dados de localização.

Visando atenuar os riscos e preservar a privacidade, seus criadores definiram algumas regras: as informações são armazenadas em um servidor de computador remoto apenas por 14 dias; somente as autoridades de saúde podem criar aplicativos a partir desse sistema; é proibido compartilhar localização de usuário, bem como usar os dados para publicidade ou policiamento; os gestores dos aplicativos são obrigados a obter o consentimento do usuário antes de usar a api de notificação de exposição; e é obrigatório um consentimento específico para compartilhar os resultados positivos dos testes com as autoridades de saúde pública.

Mesmo com essas regras, não há unanimidade: no início de maio, o Reino Unido decidiu evitar o sistema projetado pela Apple e pelo Google; a preocupação dos especialistas britânicos é quanto a sua eficácia, na medida em que depende de ações dos usuários, inclusive o de manter seus celulares sempre ligados. Além disso, como mostra estudo da Universidade de Oxford, sua eficiência depende de que ao menos 60% da população baixe o aplicativo.

Usando tecnologia de *blockchain*, o Banco Interamericano de Desenvolvimento (bid), em parceria com três empresas privadas – Everis, iovlabs e World Data –, anunciou o lançamento do aplicativo david19. A ideia é oferecer aos cidadãos latino-americanos uma plataforma para o compartilhamento de dados relacionados à covid-19, facilitando a gestão de políticas públicas. A escolha do nome David para o aplicativo é uma alusão à história bíblica bem conhecida do enfrentamento vitorioso entre o frágil Davi e o invencível gigante Golias, nesse caso o coronavírus.

Segundo seus idealizadores, o objetivo é construir um registro comum descentralizado do status de cada usuário, gerando mapas de situações de risco. O projeto é inspirado em iniciativas bem-sucedidas em Cingapura e Coreia do Sul. O bid assegura que não há chance de o anonimato ser violado.

Em outra direção, no Brasil temos o Sistema de Monitoramento Inteligente (simi), desenvolvido pelas quatro principais operadoras de telefonia – Claro, tim, Oi e Vivo – para monitorar o isolamento durante a quarentena. Por meio de “mapas de calor”, captados dos celulares, o sistema identifica aglomerações, com o objetivo de apurar o índice de isolamento social por região. As primeiras experiências foram no Rio de Janeiro, parceria com a tim, e em São Paulo,

parceria do estado com as quatro operadoras. Só os governos têm as “chaves de acesso à plataforma”. Os dados são anonimizados, ou seja, o sistema não permite identificar o cliente. É como uma catraca de metrô, quando o usuário passa, apenas um número é registrado, nenhuma informação pessoal (a apuração indica quantas pessoas passaram na catraca, e não quem passou). O simi está programado para disparar sinais de alerta via sms aos usuários.

Os habitantes de Wuhan, na China, aparente origem da covid-19, aos poucos retomam suas atividades, mas com novas rotinas. Na fábrica da Lenovo, por exemplo, os funcionários passam diariamente por quatro checagens de temperatura, sendo liberados apenas aqueles com temperaturas abaixo de 37,3 °C; a arquitetura da fábrica foi alterada, reduzindo em 50% a relação trabalhador/espço; robôs desinfetam, constantemente, as instalações e executam pequenas tarefas, reduzindo a circulação de pessoas.

Epidemiologia digital no combate à covid-19: inovando no monitoramento de epidemias

22.5.2020

O filósofo alemão Peter Sloterdijk, refletindo sobre a pandemia da covid-19, propôs o conceito de “coimunidade”, cujo significado é o compromisso individual de proteção mútua que, segundo ele, será a nova maneira de estar no mundo: “O conceito de *coimunidade* implica aspectos de solidariedade biológica e de coerência social e jurídica. Essa crise revela a necessidade de uma prática mais profunda do mutualismo, ou seja, proteção mútua generalizada”. Para o filósofo, a crise atual evidencia a extrema interdependência, o que requer uma declaração geral de dependência universal. “A imunidade será a grande questão filosófica e política após a pandemia”, vaticina Sloterdijk.¹⁸⁰

A covid-19 acelerou a necessidade de proteção mútua, consolidando uma nova forma de monitorar epidemias em tempo real, a “epidemiologia digital”. A Associação Internacional de Epidemiologia (iea) define “epidemiologia” como o estudo dos fatores que determinam a frequência e a distribuição das doenças nas coletividades humanas; a epidemiologia digital incorpora as tecnologias digitais a esses estudos.

O uso de dados gerados fora do sistema público de saúde – celulares, *wearables*, videovigilância, mídias sociais, pesquisas na internet – para monitorar epidemias não é novo; os sistemas de rastreamento digital foram usados na epidemia haitiana de cólera, em 2010, e no surto de ebola em 2014; o coronavírus ampliou a dimensão e a diversidade desses sistemas.

Na Nova Zelândia, na Tailândia e em Taiwan, por exemplo, as autoridades identificam os cidadãos que estão desrespeitando as normas vigentes pelos dados de localização de celulares, penalizando os infratores com multas, em alguns casos, de valores significativos. Na China, na Polônia e na Rússia os governos estão utilizando tecnologias de reconhecimento facial.

O governo de Cingapura, em março, solicitou aos cidadãos a instalação em seus *smartphones* do aplicativo TraceTogether: com tecnologia de *bluetooth*, o *app* troca números de identificação com os celulares de outros usuários a menos de um metro e meio de distância, compartilhando com o governo dados de usuários positivos para a covid-19. A eficiência do aplicativo está no fato de o vírus ser transmissível por meio de contato casual ocorrido dias antes dos primeiros sintomas, e dificilmente as pessoas se lembram de todos os contatos que tiveram ao longo de vários dias.

O rastreamento de contato digital via *smartphone* oferece uma alternativa eficaz e menos onerosa, mesmo que nem todo mundo tenha o dispositivo.

Pesquisas no Reino Unido sugerem que a covid-19 poderia ser eliminada se 80% dos usuários de *smartphones* (56% da população em geral, assumindo 70% de penetração de *smartphones*) usassem o aplicativo. Os resultados, contudo, dependem do nível de adesão da população, que, por sua vez, depende de campanhas de esclarecimento e da transparência sobre como os dados serão coletados e utilizados, e quais os benefícios.

Em Cingapura, por exemplo, apenas cerca de 20% dos usuários de *smartphones* instalaram o aplicativo TraceTogether, mesmo que o compartilhamento dos dados com as autoridades de saúde seja restrito a usuários infectados. Pesquisa realizada nos Estados Unidos indicou que cerca de 40% de usuários de *smartphones* disseram que com certeza optariam por um aplicativo de rastreamento de contatos, e pouco menos de 70% disseram que provavelmente o fariam; no entanto, outra pesquisa encontrou taxas de adesão muito mais baixas: 17% com certeza, 32% provavelmente.

As tecnologias são úteis para rastrear a disseminação do vírus e interromper o agente infeccioso, mas, simultaneamente, suscitam questões éticas sensíveis. Michelle Mello e Jason Wang, pesquisadores da Faculdade de Medicina da Universidade de Stanford, em artigo publicado na revista *Science*, exploram algumas preocupações, particularmente a privacidade e a autonomia.¹⁸¹ Os autores alertam que alguns desses sistemas estão ameaçando a privacidade em níveis maiores do que os usos correntes em finalidades comerciais, e nem sempre estão respeitando a autonomia de consentir ou não com o acesso às informações pessoais. Os riscos são minimizados nos sistemas que usam dados anonimizados (sem identificação individual), o que não é o caso de vários dos sistemas associados ao combate à covid-19, como o aplicativo Alipay, na China, que atribui categorias de risco aos indivíduos. Outra preocupação dos pesquisadores é quanto ao viés contido nos dados, ao não representar o universo da população em estudo (há grupos sub-representados nos dados de laboratórios e centros de saúde e no próprio acesso à internet e celulares).

Mello e Wang apontam três fatores que potencialmente amplificam os efeitos negativos. O primeiro deles seria o escopo: o uso de grandes conjuntos de dados, representando um aumento significativo do número de pessoas em análise. O segundo, a velocidade: a urgência para desenvolver e implantar aplicativos e algoritmos pode comprometer a validação dos testes e, conseqüentemente, a confiança. E por último, a fonte: os dados captados fora do sistema de saúde, por exemplo, nas mídias sociais, requer validação pela alta probabilidade de serem imprecisos ou incompletos.

A transparência dos sistemas de rastreamento pode minimizar os riscos à privacidade, já que não é aconselhado apostar na eficácia da supervisão dos órgãos públicos. No caso dos Estados Unidos, Mello e Wang sugerem que as agências federais de saúde encomendem estudos às Academias Nacionais de Ciências, Engenharia e Medicina, recomendando regras para a vigilância digital em

pandemias que incluam mudanças nas leis estaduais e federais de privacidade e poderes emergenciais.

A inteligência artificial não é inteligente, é pura estatística

5.6.2020

O telefone toca e um médico atende: “Senhor, ficamos sem ventiladores. O que fazer quando chegarem mais pacientes?”. Essa é a cena inicial do filme *Vírus*, de 2019, dirigido por Aashiq Abu e produzido em Mollywood – cinema sediado no estado indiano de Kerala e falado em malaiala –, que narra a luta para conter o surto do vírus Nipah, em 2018, naquela região.

O patógeno, transmitido por morcego, matou 21 das 23 pessoas infectadas naquele ano. A epidemia, contudo, foi vencida em um mês com as medidas adotadas pelo governo local: rastreamento de contatos e quarentena. Segundo a oms, os sintomas iniciais foram febre, dor de cabeça, mialgia (dor muscular), vômitos e dor de garganta; alguns tiveram pneumonia atípica e problemas respiratórios graves. O período de incubação variou de quatro a 14 dias. Até hoje, não existem medicamentos ou vacinas, embora a oms tenha identificado o vírus Nipah como doença prioritária. Qualquer semelhança não é mera coincidência.

Desde o primeiro alerta da canadense BlueDot, em 31 de dezembro de 2019, ao detectar uma nova forma de doença respiratória na região do mercado de Wuhan, na China, passando pelos instrumentos de rastreamento até o desenvolvimento de vacinas, os modelos de inteligência artificial baseados em dados têm contribuído no enfrentamento da epidemia. Os resultados, contudo, são comprometidos pelas ainda limitações dos modelos. Vejamos algumas abordagens a essas limitações.

Numa economia orientada por dados (*data-driven economy*), estes são um ativo estratégico que se distingue dos demais pela não rivalidade (os dados podem ser usados, simultaneamente, por vários usuários) e pela externalidade (o valor dos dados para qualquer usuário depende das ações realizadas por outros, ou seja, a aplicabilidade). A erosão da privacidade é uma externalidade negativa. Diane Coyle, professora da Universidade de Cambridge, contudo, alerta para a capacidade potencial dos dados de produzir externalidades positivas se os bancos de dados não fossem isolados em diferentes organizações públicas e privadas. A falta de integração dos dados de pacientes infectados, por exemplo, tem sido uma barreira no enfrentamento da covid-19.

No entanto, mesmo se for firmado um acordo de compartilhamento de dados, ao menos para fins extraordinários, como a epidemia da covid-19, os especialistas ainda terão de lidar com o problema do viés contido nos dados, responsável em parte pelos resultados discriminatórios dos modelos baseados na técnica de redes neurais profundas (*deep learning*). Diversos centros de pesquisa

de universidades e empresas de tecnologia estão investindo pesado em soluções para eliminar ou minimizar esse viés; a expectativa é alvissareira para os próximos anos.

O problema do viés é abordado em inúmeros estudos e publicações. A título de ilustração, seguem três exemplos: estudo da ProPublica, realizado no condado de Broward, Flórida, mostrou que o sistema compas – sistema de automatização de decisões sobre liberdade condicional adotado em vários estados norte-americanos – atribuiu aos réus afro-americanos a classificação de “alto risco” quase o dobro de vezes comparativamente aos réus brancos.¹⁸² Estudo das pesquisadoras Joy Buolamwini e Timnit Gebru apurou resultados discriminatórios com viés de gênero e raça em algoritmos de análise facial.¹⁸³ Em outro estudo, da Universidade de Washington, na busca por imagens de ceos, apenas 11% dos principais resultados mostravam mulheres, embora elas representassem 27% dos ceos dos Estados Unidos na época. Provavelmente, na base de dados de treinamento dos algoritmos de ia, os grupos estavam sub-representados. Jake Silberg e James Manyika, do McKinsey Global Institute, argumentam que, dependendo da forma como os sistemas são desenvolvidos, implementados, utilizados e interpretados, podem reduzir, perpetuar ou multiplicar o preconceito humano contido nos dados.¹⁸⁴

Segundo os autores, técnicas inovadoras de treinamento, como o *transfer learning*, têm obtido resultados positivos na redução do viés particularmente em sistemas de reconhecimento facial. No “aprendizado de transferência”, algoritmos que foram treinados para reconhecer imagens de carros, por exemplo, com ajustes podem ser utilizados no reconhecimento de caminhões, encurtando o tempo de treinamento da segunda tarefa (precondição: semelhança entre os conjuntos de dados). Trata-se de armazenar o conhecimento adquirido na resolução de uma tarefa e aplicá-lo a uma tarefa diferente, porém relacionada.

Outro problema dos sistemas de redes neurais profundas (*deep learning*) é a não explicabilidade, limitação sensível no setor de saúde. Seus profissionais resistem a adotar resultados de sistemas inteligentes sem saber os meandros do processo, ou seja, como os sistemas chegaram àquele resultado específico. No estágio de desenvolvimento atual da ia, as previsões não devem ser aceitas como definitivas (além disso, toda previsão contém margem de erro).

A inteligência artificial que está sendo usada em larga escala são modelos estatísticos de probabilidade. Esses modelos não são inteligentes, não têm capacidade de compreender o significado, não entendem o contexto social, o que se constitui em sua principal fragilidade (ou vulnerabilidade). O papel desses modelos é subsidiar o processo de tomada de decisão dos profissionais humanos.



Futuro sustentável e intensivo em tecnologia é a meta. Essa é a aposta (e o desejo), pelo menos, de uma parte da sociedade brasileira: o Google Trends indica um aumento significativo, a partir de 2019, na busca pelo termo “inteligência artificial”, e o estudo publicado pelo Google em 2021 sinaliza um aumento de 247% nas buscas por assuntos relacionados à crise climática, e 88% dos usuários consultados acreditam que as empresas/marcas devem agir para amenizar os efeitos da crise ambiental.¹⁸⁵

Existe uma sobreposição dos impactos éticos e sociais das duas agendas, por exemplo, no tema da automação e os subsequentes efeitos das tecnologias digitais sobre o trabalho (substituição do trabalhador, demanda por novas habilidades, redução da renda e aumento da desigualdade), e as inéditas interfaces homem-máquina; a Economia Verde e a Economia de Dados são intensivas em tecnologia (e não em mão de obra).

A ia é estratégica no enfrentamento das mudanças climáticas, seus modelos são usados para monitorar o aquecimento global. Entre os especialistas do clima predomina, contudo, a ideia de ia como “ferramenta”. É urgente superar essa visão e convergir as duas agendas.

O movimento por uma “ia sustentável”, além dos aspectos éticos, considera os impactos negativos dos robustos sistemas de ia sobre o meio ambiente, que é o tema central dos dois primeiros artigos deste bloco (a natureza não artificial da ia). O terceiro e último artigo debate o argumento inicial: a convergência das agendas climáticas e de ia como estratégica para pensar o futuro.

Consumo de energia e emissão de CO_2 dos algoritmos de inteligência artificial: como evitar uma catástrofe climática

18.12.2020

Estudo da Universidade de Stanford de 2019 descobriu que o treinamento de um sistema padrão de processamento de linguagem de inteligência artificial gera, aproximadamente, a mesma quantidade de emissão de carbono produzida por uma pessoa em uma viagem de ida e volta entre Nova York e São Francisco. O processo completo de construir e treinar esse sistema pode gerar, dependendo da fonte de energia, o dobro de emissão de CO_2 de um americano médio ao longo de toda a vida (com a ressalva de que não é simples medir o consumo de energia individual).

No mesmo ano, a pesquisadora de inteligência artificial da Universidade Carnegie Mellon, Emma Strubell, alertou que desde 2017 o consumo de energia e a pegada de carbono estão explodindo à medida que os modelos de ia são alimentados com mais e mais dados. O estudo de Strubell descobriu que um modelo específico de rede neural profunda (*deep learning*) teria produzido o equivalente a 626.155 libras (284 toneladas métricas) de dióxido de carbono, o que corresponde à produção vitalícia de cinco carros norte-americanos médios.

Existe uma relação direta entre o consumo de energia e a emissão de carbono (CO_2). De acordo com a Agência de Proteção Ambiental dos Estados Unidos (U.S. Environmental Protection Agency – epa), um quilowatt-hora de consumo de energia gera 0,954 libras de emissões de CO_2 em média no país, considerando, proporcionalmente, as distintas fontes de eletricidade em toda a rede de energia americana – energias renováveis, nuclear, gás natural, carvão. Segundo Strubell, a pegada de carbono seria menor se os modelos de inteligência artificial fossem todos treinados com energia renovável, e compara, por exemplo, o Google Cloud Platform e a aws/Amazon, o primeiro com matriz energética renovável.

Rob Toews, em artigo publicado na *Forbes*, estima que, entre 2012 e 2018, a quantidade de energia consumida para rodar os modelos de inteligência artificial tenha aumentado 300 mil vezes, fruto principalmente do aumento da quantidade de dados utilizada no treinamento: em 2018, o modelo bert (algoritmo do buscador do Google) alcançou o melhor desempenho da classe de processamento de linguagem natural (pnl) – técnica de ia que ajuda as máquinas a analisar, interpretar e gerar textos –, ao ser treinado em um conjunto de dados de 3 bilhões de palavras; o gpt-2 (Generative Pre-Training for Language Understanding), lançado em junho 2018 pela Openai, foi treinado em um conjunto de dados de 40 bilhões de palavras, e o gpt-3, sucedâneo do gpt-2,

lançado em junho 2020, foi treinado com cerca de 500 bilhões de palavras.¹⁸⁶

A inteligência artificial disseminada atualmente na sociedade e na economia, baseada na técnica de redes neurais profundas (*deep learning*), é um modelo empírico de experimentação e ajustes, que demanda um exercício de “tentativa e erro”, ou seja, rodar o sistema inúmeras vezes com distintas arquiteturas, hiperparâmetros e algoritmos. No modelo gpt-3 da Openai, por exemplo, ao longo de seis meses foram testadas 4.789 versões diferentes, demandando um total de 9.998 dias de tempo de gpu (*graphics processing unit*), e gerando mais de 78 mil libras de emissões de CO_2 , uma quantidade maior do que se estima que um adulto norte-americano médio produz em dois anos. Como enfatiza Rob Toews, no artigo citado, o alto consumo de energia extrapola o período de treinamento: a multinacional de tecnologia Nvidia estima que 80% a 90% do custo de uma rede neural ocorre na etapa de implementação.

Tentar controlar essa externalidade negativa, ou ao menos minimizá-la, é o propósito do movimento denominado de Green AI. Com foco inicial em conscientizar os desenvolvedores de inteligência artificial e as empresas que utilizam essas tecnologias em larga escala, os pesquisadores sugerem medidas relativamente simples, como privilegiar algoritmos que consomem menos energia e transferir o treinamento para locais remotos ou com fonte de energia renovável: um processamento na Estônia, por exemplo, que depende predominantemente do óleo de xisto, produz 30 vezes o volume de carbono que o mesmo processamento produziria em Quebec, que depende principalmente da hidroeletricidade¹⁸⁷.

Uma equipe de pesquisadores da Universidade de Stanford, do Facebook ai Research e da Universidade McGill criou um dispositivo fácil de usar para medir a quantidade de energia que um projeto de aprendizado de máquina usa e quanto isso significa em emissões de carbono. “Há um grande esforço de expandir o aprendizado de máquina para resolver problemas cada vez maiores, usando mais poder de computação e mais dados. Quando isso acontece, temos de estar atentos para saber se os benefícios desses modelos de computação pesada compensam o custo do impacto no meio ambiente”, pondera Dan Jurafsky, professor titular de Ciência da Computação em Stanford e um dos membros da equipe.¹⁸⁸

As gigantes de tecnologia estão, em parte, movimentando-se. O Google exige “emissões líquidas de carbono zero” para seus *data centers*, compensadas por extensas compras de energia renovável, e se autodenomina “nuvem mais ecológica do setor”, afirmando que 100% do consumo de energia do Google Cloud é proveniente de energia renovável. A Microsoft anunciou no início do ano um plano para se tornar “carbono negativo” até 2030. Por outro lado, os principais provedores de “nuvem” – Microsoft, Amazon, Google – não divulgam as demandas de energia de seus sistemas de aprendizado de máquina. A estratégia dessas empresas, aparentemente, está mais calcada na compra de crédito de carbono do que em mudanças efetivas em suas operações.

Dado o impacto no meio ambiente dos modelos de inteligência artificial, é crítico, no mínimo, aumentar a transparência, estabelecendo indicadores universais de mensuração, ou seja, cada novo modelo de ia desenvolvido e/ou implementado deve incluir a informação de quanta energia foi gasta no processo.

Em paralelo, a comunidade de ia – desenvolvedores, pesquisadores, instituições e empresas – precisa investigar novos paradigmas que dispensem o uso de conjunto de dados cada vez maiores, e com crescimento exponencial. Outra ponderação de Emma Strubell, diz respeito à desigualdade entre a academia e as grandes plataformas de tecnologia: “Essa tendência de treinar modelos enormes em toneladas de dados não é viável para acadêmicos, portanto, há uma questão de acesso equitativo entre pesquisadores na academia e pesquisadores na indústria”.¹⁸⁹

São imensos os benefícios possibilitados pelas tecnologias de inteligência artificial, inclusive no enfrentamento das mudanças climáticas. Estudo elaborado por 23 pesquisadores, incluindo um dos “pais” da técnica de *deep learning*, Yoshua Bengio, publicado em novembro 2019,¹⁹⁰ apresenta 13 áreas em que a ia pode colaborar, tais como produção energética, remoção de CO_2 , educação, geoengenharia solar e finanças, construções energeticamente eficientes, desenvolvimento de novos materiais com uso reduzido de carbono, além de maximizar o monitoramento do desmatamento e do transporte ecológico. O que ainda é relativamente pouco debatido é o consumo de energia e a consequente emissão de carbono desses mesmos sistemas inteligentes. O desafio é achar o ponto de equilíbrio.

Inteligência artificial não é inteligente nem artificial

14.5.2021

Não existe uma definição universal de “inteligência”. Genericamente, o termo designa a capacidade de um agente atingir objetivos determinados em uma ampla gama de domínios e, em geral, é associado à espécie humana. Definições específicas contemplam atributos intangíveis, como a capacidade de fazer analogia e entender o significado, além de consciência, intencionalidade, livre-arbítrio, ética, moral. Por esses parâmetros, no estágio de desenvolvimento atual, a inteligência artificial não é inteligente.

A ia também não é “artificial”, seu desenvolvimento e uso extrapolam a esfera abstrata dos algoritmos, dependem de infraestruturas físicas que, por sua vez, dependem de recursos naturais, particularmente o lítio. De cor prateada, maleável, denominado de “ouro branco”, o mineral é estratégico para o setor de tecnologia: os equipamentos eletrônicos portáteis (*smartphones*, *notebooks*, câmeras digitais, assistentes virtuais), interface entre os usuários e os sistemas de inteligência artificial, funcionam com baterias de íons de lítio (lib). Gradativamente, elas estão sendo implantadas nos *data centers*, em função da maior densidade de energia e vida útil significativamente mais longa, além de “pegada ecológica” (*“ecological footprint”*).

O lítio foi descoberto em Silver Peak, Nevada, Estados Unidos, durante a Segunda Guerra Mundial. Por mais de 50 anos, o mineral foi extraído em quantidades modestas, até ser valorizado no século XXI pelo setor de tecnologia. Rockwood Holdings é a única mina de lítio em operação nos Estados Unidos. Fundada em 2000, em Princeton, Nova Jersey, foi adquirida em 2014 pela fabricante de produtos químicos Albemarle Corporation por 6,2 bilhões de dólares. A maior reserva de lítio do mundo, contudo, está no deserto do Salar de Uyuni, no sul da Bolívia, que, junto com o Salar de Atacama, Chile, e o Salar del Hombre Muerto, Argentina, formam o “triângulo do lítio”, representando cerca de 68% das reservas mundiais. As demais regiões com minas de lítio são Congo, Mongólia, Indonésia e desertos da Austrália Ocidental. As condições de trabalho nessas minas são precárias; a Intel e a Apple, recentemente, foram criticadas por auditar apenas as fundições (não as minas) na busca do Certification of Conflict-free Status (Certificação de Status Livre de Conflito).

As baterias de lítio são, igualmente, utilizadas nos carros elétricos, que incluem os carros autônomos. A maior fábrica de baterias de lítio do mundo é a Tesla Gigafactory, projeto da Tesla com a Panasonic, em Nevada, EUA; na produção de veículos elétricos, a Tesla consome cerca de 50% do total de baterias de lítio do mundo (anualmente, cerca de 28 mil toneladas de hidróxido de lítio).

A mineração, fundição, exportação, montagem e transporte da cadeia de abastecimento da bateria têm fortes impactos negativos no meio ambiente e nas comunidades locais.

Kate Crawford – cofundadora do instituto de pesquisa ai Now, da Universidade de Nova York, professora da usc Annenberg, de Los Angeles, e professora-visitante de ia e Justiça na École Normale Supérieure, em Paris, além de pesquisadora sênior da Microsoft Research – lançou o livro *Atlas of AI*.¹⁹¹ Com base em minuciosa e extensa pesquisa de campo, Crawford alerta para o fato de que as externalidades negativas da inteligência artificial transcendem as questões éticas, na medida em que produzem mudanças geomórficas profundas e duradouras ao planeta, destacando o protagonismo do lítio, cada vez mais escasso, e da energia, ainda fortemente concentrada em energias fósseis.

Crawford denuncia o “mito da tecnologia limpa” das grandes empresas de tecnologia, particularmente das plataformas de nuvem – Amazon Web Services, Microsoft Azure, Google Cloud e ibm Cloud: divulgam políticas ambientais, iniciativas de sustentabilidade e contribuições da ia no enfrentamento das questões climáticas e escondem como segredos corporativos os danos, entre outros, da extração de lítio e do consumo de quantidades gigantescas de energia. Google e Apple se declararam “carbono *free*”, ou seja, compensam suas emissões de carbono comprando crédito, o que indica a incapacidade de reduzir suas próprias emissões.

O consumo de energia, e consequente emissão de CO_2 , tem crescido significativamente com o aumento exponencial da velocidade e da precisão dos modelos de inteligência artificial. Emma Strubell, pesquisadora de ia da Universidade de Massachusetts Amherst, estimou que, para rodar apenas um modelo de *natural language processing* (nlp), são produzidas mais de 660 mil libras de emissões de dióxido de carbono, o equivalente a cinco carros movidos a gasolina durante sua vida útil total (incluindo o período de fabricação), ou 125 voos de ida e volta de Nova York a Pequim.¹⁹²

A Openai, instituição de pesquisa sem fins lucrativos dedicada a promover a inteligência artificial amigável, fundada por Elon Musk e outros investidores do Vale do Silício, estima que desde 2012 a quantidade de computação usada para treinar um único modelo de ia tem aumentado por um fator de 10 a cada ano.¹⁹³ Os *data centers* são grandes poluidores. Na China, por exemplo, 73% da energia utilizada nos centros de processamento de dados vem do carvão; em 2018, eles emitiram cerca de 99 milhões de toneladas de CO_2 , e a previsão é aumentar em dois terços até 2023.¹⁹⁴

A inteligência artificial hoje não é inteligente, não é artificial, nem objetiva e neutra. Como pondera Kate Crawford, os sistemas de ia “estão embutidos nos mundos social, político, cultural e econômico, moldados por humanos, instituições e imperativos que determinam o que eles fazem e como o fazem”.

A exposição *Futures*, inaugurada em novembro no Arts and Industries Building, em Washington, e projetada pelo premiado escritório de arquitetura Rockwell Group, especula, com arte e tecnologia, o próximo capítulo da humanidade: “Sinta o cheiro de uma molécula. Limpe suas roupas em um pântano. Medite com um robô de ia. Viaje no espaço e no tempo”, anuncia a curadoria. Um dos cenários é o *Futures that Work*, que apresenta, entre outras invenções, o “Virgin Hyperloop”, tubo de vácuo para transporte de passageiros que encurta distâncias de horas para minutos; o “Ecos BioReactor”, reator que captura carbono do ar e converte em biomassa; o “Mineral Rover”, sistema que provê a quantidade de minerais e água necessários para cada planta (personalização); e o “Waha Water Harvester”, equipamento, movido a energia solar, que puxa água diretamente do ar. São “futuros” sustentáveis e intensivos em tecnologia.

A simples substituição de viagens de negócio por videoconferências tem efeito benéfico sobre a pegada de carbono, ilustrando a interdependência entre sustentabilidade e tecnologia. Esse exemplo consta do documento “Shaping Europe’s Digital Future”, lançado pela Comissão Europeia em fevereiro de 2020.¹⁹⁵ O documento enfatiza a relevância das tecnologias digitais para o Acordo Verde Europeu¹⁹⁶ e para os Objetivos de Desenvolvimento Sustentável,¹⁹⁷ ou seja, para a economia circular, para a descarbonização e redução da pegada ambiental e social dos produtos, particularmente em setores-chave como agricultura de precisão, infraestruturas inteligentes de transporte e energia. Para Ursula von der Leyen, presidente da comissão, trata-se de um duplo desafio: transição verde e digital.

Alinhado com essa tendência, o documento “Clima e desenvolvimento: visões para o Brasil 2030”¹⁹⁸ lançado no espaço Brazil Climate Action Hub, na Cop 26, realizada entre 31 de outubro e 12 de novembro deste ano, em Glasgow, na Escócia – tem como principal objetivo propor caminhos de transição do modelo atual de desenvolvimento do Brasil para um modelo de zero emissões líquidas de carbono. O foco do documento, do ponto de vista de ações a serem implementadas, contempla monitorar o desmatamento, promover o crescimento econômico, incentivar a agricultura sustentável e otimizar e diversificar os modos de transporte de carga e transporte público, metas viabilizadas, em parte, por modelos de inteligência artificial. Segundo Natalie Unterstell, presidente do Instituto Talanoa, responsável pelo documento ao lado do Centro Clima da ufrrj,

a abrangência desses objetivos significa, na prática, refundar as bases da economia: “Não estamos propondo uma transição só de fontes energéticas. Mas também uma transição de empregos, uma transição de modelos de negócio, uma transição de modelo de país. É disso que estamos falando”.¹⁹⁹

O artigo “The Role of Artificial Intelligence in Achieving the Sustainable Development Goals”, publicado em 2020 na revista *Nature Communications*, aborda as contribuições da inteligência artificial para o cumprimento dos 17 objetivos e 169 metas de desenvolvimento sustentável da Agenda 2030 da Onu.²⁰⁰ Com 10 autores, entre eles Virginia Dignum, da Universidade de Umeå, e Max Tegmark, do MIT, o artigo analisa, com base em pesquisas empíricas, os impactos positivos e negativos da IA em três pilares – sociedade, economia e ambiente. No pilar “ambiente”, o estudo identificou 25 alvos, 93% do total, para os quais a inteligência artificial poderia atuar como um facilitador, com destaque para a compreensão e a modelagem dos possíveis impactos das mudanças climáticas.

As tecnologias de inteligência artificial permitem extrair informações úteis de grandes conjuntos de dados gerados por sensores alocados estrategicamente e, com base neles, entre outras ações, alimentar os sistemas de monitoramento de desmatamento e degelo e identificar as correlações entre regiões aparentemente não correlacionadas; por exemplo, os efeitos dos eventos no Saara sobre a Amazônia (muito bem retratados no documentário *Mundo conectado*, da Netflix). O uso mais eficiente dos recursos naturais passa, igualmente, pelas tecnologias de inteligência artificial, assim como as chamadas *smart sustainable cities* (monitoramento da mobilidade urbana, gerenciamento de tráfego, gestão eficaz das infraestruturas e serviços como energia e saneamento básico).

A plena concretização da agenda climática depende do uso intensivo de tecnologias de inteligência artificial, com impactos éticos e sociais. No âmbito social, destacam-se os efeitos diretos e indiretos sobre o trabalho: deslocamento do trabalhador com a substituição por automação inteligente em um conjunto crescente de funções; efeito negativo sobre a renda, ao aumentar a competição pelas funções remanescentes preservadas, por enquanto, aos humanos, e mudança na interface homem-máquina em larga escala, demandando qualificação e requalificação, o que remete à formação e à educação (e não ao simples treinamento). Na próxima década, uma parcela não desprezível das novas funções será em ocupações totalmente novas ou ocupações existentes com conteúdos e requisitos de competências transformados, entre outras, pela Economia Verde e pela Economia de Dados. Igualmente em comum, emergem várias questões éticas, como privacidade e transparência no uso de dados pessoais e nas decisões automatizadas.

A sociedade encontra-se em um ponto crítico das duas agendas – mudanças climáticas e tecnologias de inteligência artificial –, ambas estratégicas para garantir um futuro sustentável para a humanidade. Atualmente, a IA é pouco considerada nas reflexões éticas e sociais dos pensadores das questões climáticas;

prevalece uma abordagem instrumental da ia. É urgente mudar essa visão e estabelecer uma sólida aliança entre as duas agendas.



IA e cultura

Na introdução ao livro *ai 2041: Ten Visions for Our Future* (2021), em coautoria com Kai-Fu Lee, Chen Qiufan – premiado autor, produtor e curador, e presidente da World Chinese Science Fiction Association – descreve o impacto sobre ele causada pela exposição *ai: More Than Human*, coproduzida por Barbican International Enterprises em colaboração com Groninger Forum: “Como uma refrescante chuva de verão, a exposição clareou meus sentidos – e mudou a maioria dos meus preconceitos preexistentes e equívocos em relação à inteligência artificial”.

Amplamente reverenciada pelo público e pela mídia, a exposição mostrou, com diversidade e complexidade, a linha do tempo representativa da evolução da ia, começando pelo personagem místico do folclore judaico Golem, criado em 1580 pelo grão-rabino de Praga Judah Loew, e o herói do anime japonês *Doraemon* (*O Gato do Futuro*), criado em 1969 e em 2002 aclamado como herói asiático pela revista *Time Asia*. São apresentados projetos proeminentes de instituições como DeepMind, Jigsaw, mit Computer Science and Artificial Intelligence Laboratory, Sony Computer Science Laboratories, e artistas em colaboração com cientistas-pesquisadores, como Joy Buolamwini, Mario Klingemann, Kode 9, Lawrence Lek, Massive Attack, Lauren McCarthy, Yoichi Ochiai, Neri Oxman, Anna Ridler, Chris Salter, Sam Twidale e Marija Avramovic e Universal Everything. Participa, igualmente, Hiroshi Ishiguro, criador de uma réplica robótica de si mesmo.

São múltiplos os movimentos artísticos baseados em ia – ultra fractal, arte genética, proceduralismo e arte transumanista –, e os sites dedicados a esses artistas – The Algorithms, Algorithmic Worlds, The Art. A aplicação das tecnologias de ia na arte extrapola a criação, sendo usada, por exemplo, no reconhecimento de autenticidade de obras de arte, no resgate de obras danificadas, na disseminação de conteúdo, na preservação de patrimônio e na experiência do visitante em museus e exposições. O uso de ia nesses domínios requer estreita colaboração entre cientistas/desenvolvedores e artistas, aproximando universos distintos (lógica, linguagem, metodologia).

Este bloco tem apenas dois artigos: o primeiro, dos primórdios da coluna, pondera sobre a viabilidade de criar arte com ia; o segundo, mais recente, traz uma visão geral das possibilidades da colaboração ia e arte.

A casa de leilões Christie's, fundada em 1766 por James Christie, é uma conceituada instituição britânica ligada ao comércio de arte. Em outubro de 2018, sua filial em Nova York leiloou uma pintura criada inteiramente por algoritmos de inteligência artificial, conquistando visibilidade na mídia mundo afora: o retrato de Edmond de Belamy, interpretando um cavalheiro aristocrático. O valor inicial foi fixado entre 7 mil e 10 mil dólares, tendo sido vendida por 433 mil dólares.

A obra pertence a uma série de imagens chamada *La famille de Belamy*, criada pelo Obvious, coletivo de artistas e pesquisadores de ia baseado em Paris. Seu propósito é explorar a interface entre a arte e a ia utilizando a arquitetura de redes neurais gan (*generative adversarial network*, rede adversária generativa).

Richard Lloyd, da Christie's, justificou a escolha da obra pela limitada intervenção humana em seu processo criativo: "O Obvious tentou limitar a intervenção humana tanto quanto possível, de modo que o trabalho resultante refletisse a forma 'purista' da criatividade expressa pela máquina".²⁰¹

A gan, introduzida em 2014 por pesquisadores da Universidade de Montreal, é uma arquitetura de redes neurais (*deep learning*) composta por duas redes – uma contra a outra, daí o "adversária" do nome –, treinadas para criar objetos semelhantes em qualquer domínio (música, fala, imagens, textos). Pode ser utilizada, por exemplo, no "envelhecimento" computacional para ajudar a identificar e/ou localizar pessoas desaparecidas, e na elaboração de "retrato falado" na busca de suspeitos.

A gan é capaz de gerar novas imagens, com aparência de autênticas, a partir do conjunto de imagens usadas no treinamento do sistema; no caso da obra leiloadada, o sistema foi alimentado com um conjunto de dados de 15 mil retratos pintados entre os séculos xiv e xx. O surpreendente, contudo, é que o retrato de Edmond de Belamy incorporou elementos contemporâneos, diferindo da concepção de retrato da época; esse efeito, segundo seus idealizadores, decorre de distorções do modelo de ia.

Aparentemente, as primeiras experimentações de arte com ia ocorreram em 2015 com o *software* DeepDream do Google, os resultados, contudo, foram obras estética e conceitualmente limitadas, não atraindo a atenção da crítica nem do público. O leilão da Christie's estimulou novas experimentações, inseridas num movimento artístico batizado pelo Obvious de "gan-ism", e muita polêmica. Vários artistas, utilizadores da inteligência artificial, contestam a originalidade não

apenas dessa obra, mas de todo o trabalho do Obvious, referindo-se ao coletivo mais como profissionais de marketing do que propriamente artistas.

O jornalista norte-americano Mark Anderson, num artigo para a revista *Wired* de dezembro de 2001, atribui ao singularista Ray Kurzweil o pioneirismo na associação computação e arte ao patrocinar os estudos de Harold Cohen. Como pesquisador visitante no Laboratório de Inteligência Artificial da Universidade de Stanford, em 1973, Cohen desenvolveu um programa de criação de arte chamado aaron, capaz de desenhar e pintar naturezas-mortas estilizadas e retratos com base em programa de computador (algumas de suas obras estão em coleções de museus importantes). Pamela McCorduck, no livro *Aaron's Code*, de 1991, considerou Cohen como o pioneiro de uma nova geração de criadores de estética, ou “meta-artistas”.

The Painting Fool, de Simon Colton, da Universidade de Londres, é outro sistema de ia relacionado à arte que vale a pena conhecer e acompanhar; assim como a plataforma AIArtists.org, espécie de curadoria de obras de pioneiros na arte de ia e fórum de reflexão sobre como a ia pode expandir a criatividade humana e contribuir para o entendimento da imaginação coletiva, e como podemos estabelecer parcerias criativas entre ia e humanos.

Pode parecer inusitada, até meio insólita, a automação da arte. Arte, associada à abstração e à subjetividade, soa como antítese de computador, lógico e objetivo. O fato é que proliferam tipos de arte baseadas em algoritmos – ultra fractal, arte genética, proceduralismo e arte transumanista –, e sites dedicados a esses artistas – The Algorithms, Algorithmic Worlds, The Art.

Estamos nos primórdios da ia, espera-se uma extraordinária evolução nas próximas décadas. Por enquanto fica a pergunta: a arte de inteligência artificial é capaz de nos emocionar como a arte humana?

Avanços e diversidade na interface arte e tecnologia: criatividade, inteligência e consciência

18.2.2022

O documentário *The Beatles: Get Back* (Disney+, 2021) de Peter Jackson narra a gravação, em janeiro de 1970, do penúltimo álbum da banda e do documentário *Let It Be* (os integrantes se separaram em setembro do mesmo ano). O projeto restaurou sessenta horas de filmagem a partir de um áudio captado por um único microfone, incluindo conversas paralelas e ruídos de fundo. Para decompor o som original e isolar com precisão cada faixa, a equipe de Jackson utilizou um sistema de inteligência artificial (IA) apelidado de “MAL”, uma dupla homenagem ao computador HAL, do filme *2001: uma odisseia no espaço*, e ao assistente dos Beatles, Mal Evans.²⁰²

Na música clássica, a IA permitiu um feito histórico: como parte das comemorações dos 250 anos de Beethoven (2019), uma equipe de especialistas liderada por Ahmed Elgammal, diretor do Art & AI Lab da Universidade de Rutgers, e Matthias Röder, diretor do Instituto Karajan de Salzburgo, apoiada em modelos de IA, dedicou-se a completar a 10ª Sinfonia inacabada de Beethoven (com sua morte em 1827, o compositor deixou apenas notas e rascunhos). O feito havia sido tentado por Barry Cooper em 1988, mas o musicólogo só conseguiu completar o primeiro movimento a partir de 250 compassos. A comprovação do resultado está na execução da obra final por uma orquestra.²⁰³

Em janeiro de 2022, o artista norte-americano Brian Donnelly, conhecido profissionalmente como KAWS, inaugurou a exposição *New Fiction*, simultaneamente, na Serpentine Gallery, em Kensington Gardens, Londres, e no *Fortnite*. KAWS e a Epic Games, dona desse jogo eletrônico, cocriaram a exposição digital com o Alliance Studios e a Beyond Creative. No empenho em transformar o jogo em um “destino de entretenimento social”, segundo o seu diretor Kevin Durkin, cerca de 400 milhões de usuários viram a exposição virtual.

O documentário dos Beatles, a finalização da sinfonia de Beethoven e a exposição de arte no *Fortnite* ilustram a atual diversidade da interface entre arte e tecnologia. A produção artística ocorre em uma multiplicidade de formatos e contextos criativos denominados arte computacional, arte interativa, arte eletrônica, ciberarte, arte digital, *media art*, arte generativa, entre outros. São múltiplos, igualmente, os movimentos artísticos baseados em IA como Ultra Fractal, Arte Cinética, Proceduralismo e Arte Transumanista, e os sites dedicados a esses artistas, como The Algorithms, Algorithmic Worlds e The Art. A plataforma AIArtists.org agrega a maior comunidade global de artistas envolvidos

com IA.

O Alan Turing Institute (ATI)²⁰⁴ pondera que um número crescente de artistas está experimentando a IA no aprimoramento, na simulação ou na réplica de suas criatividades. A complexidade desses sistemas requer intensa colaboração entre especialistas de IA e artistas, com o desafio de aproximar linguagens e lógicas distintas e, posteriormente, definir a quem cabe a propriedade intelectual. A IA oferece oportunidades sem precedentes para novas interfaces homem-máquina que afetam a sociedade cultural e socialmente (para o bem e para o mal).

Como a IA interage com a criatividade, especialmente em produções artísticas “autônomas”, é um tema filosófico em aberto: obras de arte produzidas por IA são criativas? Na inevitável comparação entre cérebro biológico e sistemas de IA (particularmente as redes neurais profundas, inspiradas no conceito de conexismo), a questão da criatividade remete à inteligência e à consciência, temas abordados no novo livro *Reality+: Virtual Worlds and the Problems of Philosophy*, do filósofo David Chalmers.²⁰⁵

A tecnofilosofia de Chalmers estabelece uma interação bidirecional entre filosofia e tecnologia: fazer perguntas filosóficas sobre tecnologia e usar a tecnologia para responder às perguntas filosóficas. O filósofo denomina o desafio de explicar o “problema difícil da consciência” (*hard problem of consciousness*) em contraste com os problemas fáceis de explicar, por exemplo, a inteligência que, para ele, é uma característica objetiva expressa no comportamento, enquanto a consciência é subjetiva. Chalmers levanta várias hipóteses sobre o que seria e como se comporta a consciência, apostando na viabilidade da ideia de “máquina consciente” por meio do *upload* da mente humana para um computador em um processo gradual (1% dos neurônios por dia) capaz de transferir a consciência original do cérebro humano para o cérebro maquínico: “Fazer um novo cérebro de uma só vez pode criar uma nova pessoa, mas a substituição gradual deixa a pessoa velha intacta. [...] Prefiro fazê-lo por *upload* gradual. Essa parece ser a melhor aposta para sobreviver ao processo e sair consciente do outro lado”, conclui o filósofo. Essa ideia, contudo, não parece factível frente à complexidade do cérebro humano.

Roberto Lent, professor emérito da Universidade Federal do Rio de Janeiro (UFRJ) e um dos maiores especialistas em cérebro, pondera que o tema da consciência está envolto numa nuvem de obscuridade difícil de ultrapassar: “As propriedades mentais são sempre inesperadas porque estão sujeitas à propriedade intrínseca do cérebro, que é mudar a todo momento”.²⁰⁶ Essa habilidade chama-se neuroplasticidade: a capacidade do cérebro de mudar, adaptar-se e moldar-se a nível estrutural e funcional quando sujeito a novas experiências do ambiente interno e externo. A dinâmica do cérebro é altamente modulável, e não uma cadeia de informação linear que leva diretamente a um resultado previsível.

Cada cérebro humano contém 86 bilhões de neurônios que se conectam

uns aos outros por meio de sinapses, e cada neurônio realiza 100 mil sinapses, gerando cerca de 8,6 quatrilhões de circuitos. No processo de transmissão da informação (comunicação), o cérebro conta ainda com 85 bilhões de células coadjuvantes. A cada novo aprendizado, por exemplo quando uma criança começa a aprender a ler e escrever, o hemisfério A, que originalmente se ocupava da função “reconhecimento de faces”, transfere-se para o hemisfério B, permitindo que aquele assuma as novas funções (aprender a ler e escrever não são habilidades inatas, precisam ser aprendidas).²⁰⁷

Do ponto de vista da neurociência, ainda segundo Lent, pode-se abordar o cérebro em três níveis heurísticos: a) microescala, perspectiva reducionista em que se considera os neurotransmissores, as moléculas transmissoras que participam da sinapse; b) mesoescala, com distintas graduações em que o foco são as redes de neurônios em comunicação – neurônios inibidores, excitadores e modulares – que transformam completamente a informação inicial; c) macroescala que contempla o cérebro como um todo em comunicação com outros cérebros ativando as redes de linguagem, consultando a memória em busca de nexos no diálogo, acessando circunstâncias emocionais que modificam a memória, a informação e a forma de se comunicar com os interlocutores; e d) grupos de cérebros que se comunicam com outros grupos de cérebros, por exemplo, alunos de uma sala de aula em interação simultânea com alunos de outra sala de aula.

Algumas perguntas básicas para Chalmers (ou quem estiver apto a responder): como conciliar essa complexidade com o *upload* de 1% dos neurônios por dia? Quais tipos de neurônios estariam representados nesse conjunto de 1%? Como seria feita a seleção dos neurônios? Após transferidos, como seriam conservados? Como se daria a agregação de cada conjunto de 1% de neurônios transferidos? E quais são os indicadores de que os neurônios transferidos para o computador serão substituídos por neurônios com a mesma identidade, garantindo a manutenção intacta da consciência original?

Pensar, ou mesmo meramente especular, sobre qualquer tema é válido, contribui para formular hipóteses a serem ou não comprovadas posteriormente; o risco está em construir conceitos baseados em concepções incertas, ou não claramente comprovadas por evidências científicas, por exemplo, supor a viabilidade de uma IA consciente se ainda estamos longe de dominar o conhecimento sobre o cérebro e a consciência.



Interação humano-máquina

Existe um paradoxo em nosso relacionamento com os sistemas maquínicos: a tendência inicial é confiar plenamente neles como “mestres onipotentes”, contudo, assim que descobrimos suas limitações, ou seja, que podem cometer erros, entramos num processo de rejeição absoluta, privilegiando nossa própria capacidade de julgamento, ignorando as falhas humanas, os preconceitos, as subjetividades, o inconsciente. Os especialistas denominam esse comportamento de “aversão a algoritmos”. Em geral, percebe-se uma tolerância menor aos erros dos algoritmos do que aos erros humanos. O “melhor dos mundos” é promover a parceria entre sistemas automatizados (ia) e especialistas humanos, cada um contribuindo com seus atributos distintivos.

Este bloco é o maior, com nove artigos, numa tentativa de tangenciar a profundidade e complexidade dessa inédita interface homem-máquina. São abordados, entre outros, temas como o efeito da ia sobre nossa capacidade decisória; o vício em tecnologia, tema retomado no último artigo do bloco “ia e saúde”; algumas distinções entre a inteligência humana e a inteligência de máquina, inclusive a ponderação sobre a aplicabilidade do termo; a transição do Antropoceno para o Novaceno, segundo James Lovelock, e o *Fortnite* como rede social e futuro metaverso.

Os algoritmos de inteligência artificial estão afetando nossa capacidade de decisão?

6.9.2019

O filósofo René Descartes (1596-1650) associava o animal à máquina por meio do conceito de “*bête-machine*”, estabelecendo analogias e diferenças entre o animal, o humano e o autômato. Para o filósofo, se houvessem máquinas em tudo semelhantes a um macaco, por exemplo, não teríamos meios de atestar com certeza que não seriam da mesma natureza; contudo, se houvessem outras máquinas que imitassem os humanos, sempre teríamos meios de reconhecer que não seriam verdadeiros humanos. A razão é o que distinguiria o humano do animal e da máquina, a capacidade de raciocinar (a linguagem seria outro diferencial).

Gostamos de pensar que temos controle sobre nossas decisões, o chamado “livre-arbítrio”. Será essa crença verdadeira ou devemos aceitar que as nossas decisões sempre foram influenciadas (ou moldadas/manipuladas) por todo um universo de conexões? Será que o *big data* e os novos modelos estatísticos de previsão (inteligência artificial/*machine learning/deep learning*) representam inéditos desafios à ideia-chave do Iluminismo sobre nossa autonomia e nosso livre-arbítrio?

Cada modelo econômico tem seus mecanismos de persuasão com impactos culturais e comportamentais, ou seja, transcendem o consumo. Na economia industrial, caracterizada pela produção e pelo consumo massivo de bens e serviços, a propaganda predominou como meio de convencimento e influência nas escolhas e preferências dos indivíduos (e, igualmente, no comportamento e na cultura).

Na economia da informação em rede, temos a customização de produtos e serviços, com a correspondente comunicação segmentada. A internet ampliou a intervenção externa nos processos de decisão, seja pelas redes sociais, seja pelos sofisticados mecanismos desenvolvidos pelas empresas a partir da Web 2.0. Os sistemas de cadastro das plataformas *online* (comércio eletrônico, redes sociais), por exemplo, permitiram conhecer os hábitos de consumo dos usuários, informações úteis para a elaboração de planos de marketing eficientes (vender, conquistar ou fidelizar o consumidor). Os links patrocinados do Google produziam (e produzem) resultados afirmativos que correlacionam o perfil do potencial consumidor com os atributos do produto ou serviço ofertado.

Na Economia de Dados do século xxi, a personalização (não confundir com individualização) está na base da mediação tanto de bens quanto de informação; os algoritmos de inteligência artificial promovem estratégias de comunicação mais

assertivas a partir do conhecimento captado, minerado e analisado de dados pessoais gerados nas interações no ambiente digital.

Os algoritmos de inteligência artificial personalizam as consultas ao Google, com respostas que variam em função do perfil de quem está buscando a informação; essa é uma pequena ilustração da interferência dos algoritmos no acesso à informação e, possivelmente, em nossas decisões e ações. Os tradicionais mediadores humanos estão sendo substituídos por mediadores automatizados, especificamente os algoritmos de inteligência artificial.

A rede social Facebook, no empenho de monetizar sua gigantesca base de dados, anuncia que seus algoritmos de ia são capazes de mapear a personalidade dos usuários com 80% de precisão com base nos cliques e nas curtidas, contemplando atributos como gênero, idade, formação, etnia, “desvios” de personalidade, orientação sexual, política e religiosa, doenças, uso de substâncias. De posse desse suposto conhecimento sobre seus usuários, a rede social vende aos anunciantes uma potencial comunicação hipersegmentada/personalizada de seus produtos e serviços.

Os algoritmos de ia, com base nas informações coletadas das preferências dos indivíduos, não são apenas instrumentos comerciais para ampliar vendas, muito menos circunscritos a boas intenções, como mudar, positivamente, hábitos de saúde. Esses modelos permitem prever e interferir em nossa conduta em todas as esferas da vida social, e de maneira inédita, particularmente sobre os usuários mais suscetíveis. Ainda não somos capazes de detectar a dimensão e o alcance desse processo.

E fica posto o debate: será que em algum momento tivemos realmente livre-arbítrio, ou sempre tomamos decisões influenciados por terceiros e contextos externos? Existe pensamento individual, ou é sempre formulação coletiva? Os algoritmos de inteligência artificial estão alterando a natureza dessa dinâmica, ou são simplesmente novos mecanismos de persuasão e manipulação?

Estamos viciados em tecnologia ou apenas em transição para o futuro?

20.9.2019

Em 1909, o escritor britânico E. M. Forster publicou a novela *A máquina parou*, que retrata um cenário futurista distópico, no qual somos, simultaneamente, servidos e controlados pela Máquina. Numa surpreendente antecipação tecnológica, os habitantes do planeta Terra (ano indefinido) se comunicam por meio de uma placa redonda, segurada pelas mãos, de onde emerge uma luz azulada projetando à distância imagem e som.

Construída pelo homem, a Máquina aos poucos se torna onipresente. Nada acontece fora de seu domínio, toda ação humana – da mais simples, como levantar um objeto do chão, até os relacionamentos interpessoais – é mediada pela Máquina. Vashti, palestrante plenamente integrada ao sistema, tem cinco filhos, um dos quais, Kuno, é um rebelde que percebe a eminente falência do sistema com a “parada” da Máquina. Aflito, Kuno alerta sua mãe: “Você não percebe, não percebem todos vocês, palestrantes, que somos nós que estamos morrendo, e que aqui embaixo a única coisa que de fato vive é a Máquina? Criamos a Máquina para que fizesse nossas vontades, mas agora já não podemos fazer com que atenda nossos desejos. Ela nos roubou o sentido do espaço e o sentido do tato, borrou todo tipo de relação humana e reduziu o amor a um ato carnal, paralisou nossos corpos e nossas vontades e agora nos obriga a idolatrá-la. A Máquina se desenvolve – mas não rumo a nosso destino. A Máquina segue adiante – mas não com o mesmo objetivo nosso. Existimos apenas como os corpúsculos de sangue que correm por suas artérias, e se ela pudesse viver sem nós, nos deixaria morrer”.

Numa visão para alguns utópica, em 2018, Brad Smith e Harry Shum, respectivamente presidente e vice-presidente da Microsoft, no prefácio ao livro *The Future Computed: Artificial Intelligence and its Role in Society*²⁰⁸ (*O futuro computado: inteligência artificial e seu papel na sociedade*), projetaram um cenário factível para 2038: “enquanto você dorme, seu assistente virtual – Cortana, Siri, Alexa – estará conectado com os demais dispositivos virtuais da sua casa inteligente e garantirá um despertar sincronizado, oferecendo um café da manhã ao seu gosto. Enquanto você se prepara, o assistente virtual lê as notícias, relatórios de pesquisa, as mídias sociais, destacando os temas e eventos relacionados aos seus interesses; além disso, organizará sua agenda profissional e pessoal do dia. Não haverá risco de escolher uma roupa inadequada, a previsão do tempo será certa, bem como as condições do trânsito. O seu assistente virtual igualmente lembrará dos aniversários, compromissos sociais, reservando, quando

for o caso, o restaurante de sua preferência. Avançando mais ainda, nem será necessário sair para o trabalho, você poderá realizar as reuniões em casa, alocando todos os participantes numa sala virtual movida por realidade mista (com tradução automática e simultânea para o idioma nativo de cada participante). Caso seja necessário se deslocar, um carro autônomo (sem motorista) te conduzirá, possibilitando que você trabalhe ou faça qualquer outra atividade no percurso. Os assistentes virtuais controlarão nossa alimentação, consultas e exames médicos, aprendizado, e muito mais”.

As ações da sociedade – instituições, governos, empresas, indivíduos – vão definir se chegaremos lá mais próximos da previsão de Forster ou de Smith e Shum. Por enquanto, convém ponderar sobre nossa relação atual com a tecnologia.

A Intelligent Systems Conference, realizada em Londres no início de setembro, teve como palestrante inaugural Elizabeth Churchill, diretora de experiência do usuário no Google e presidente da Association for Computing Machinery (acm). Tendo como campo de estudo a interação humano-computador, a ciência cognitiva e a psicologia experimental, nos últimos 20 anos Churchill vem se dedicando a criar aplicativos e serviços inovadores para o usuário final, particularmente nas áreas de computação ubíqua e móvel, mídia social, comunicação mediada por computador, mídia local e ciências da internet/web, no Reino Unido, nos Estados Unidos e na Ásia.

Sua motivação é identificar como os fatores técnicos, culturais e sociais afetam a maneira como as pessoas se comunicam (ou não) e colaboram (ou não) umas com as outras. Churchill reconhece que o design das plataformas digitais, particularmente das gigantes de tecnologia, é concebido para “fazer a gente voltar sempre; as empresas de internet são organizadas em torno de ferramentas de dosagem de dopamina projetadas para atrair o público”. Para ela, nossa crescente dependência das tecnologias está comprometendo a ideia de bem-estar, questão sensível a um terço dos usuários norte-americanos, preocupados com os efeitos negativos na saúde, na aptidão mental e na felicidade.

Suas pesquisas indicam que o “vício tecnológico” afeta as capacidades cognitivas dos usuários: pensamento analítico, memória, foco, reflexão sobre criatividade e resiliência mental. Enfática, a diretora do Google também alertou que a estrutura da internet e o ritmo das mudanças digitais ameaçam as interações humanas, a segurança, a democracia, os empregos e a privacidade. Churchill admite, contudo, que uma pluralidade de especialistas está convencida de que a vida digital continuará a expandir os limites e as oportunidades na próxima década, e que os benefícios serão maiores do que os malefícios.

Consciência distingue a inteligência humana da inteligência de máquina?

15.11.2019

O filme *Ela*, escrito, dirigido e produzido por Spike Jonze e lançado em 2013, narra a relação amorosa entre um escritor de cartas encomendadas pelos usuários de uma plataforma, interpretado por Joaquin Phoenix, e um sistema operacional de voz feminina, Samantha, dublada por Scarlett Johansson. Programada como uma “inteligência artificial perfeita”, que se aperfeiçoa na interação com os humanos e com os outros sistemas operacionais, Samantha simula pensamentos e sentimentos semelhantes aos humanos. A trama evolui para uma “relação de casal” entre os dois personagens, até o dia em que, por motivo não explicitado, todos os sistemas operacionais desaparecem, inclusive Samantha. No filme, Samantha interage, simultaneamente, com 8.316 usuários, sendo 641 declaradamente apaixonados.

Se Oscar Wilde estiver correto em sua famosa frase “A vida imita a arte muito mais do que a arte imita a vida” (*“Life imitates art far more than art imitates life”*),²⁰⁹ o *chatbot* Xiaoice é um belo exemplo. Criado em 2014 pela Microsoft para o mercado asiático como uma garota de 18 anos, confiável, amigável e com senso de humor, Xiaoice se relaciona regularmente com mais de 660 milhões de pessoas, sendo 100 milhões de apaixonados. Nem Samantha, pura ficção, nem Xiaoice, pura tecnologia, são dotadas de consciência.

Consciência é um termo genérico para designar uma ampla variedade de fenômenos mentais. Segundo a *Stanford Encyclopedia of Philosophy*, um sistema cognitivo pode ser considerado consciente se for capaz de sentir e responder ao seu mundo (seriam os peixes, os camarões e as abelhas conscientes?), se estiver acordado e alerta, se, além de estar consciente, também estiver consciente de que está consciente (autoconsciência, o que exclui muitos animais e crianças até certa idade). Outra forma de definir consciência é associá-la à intencionalidade (mas o que seria intencionalidade?).

Estritamente do ponto de vista científico, o que ocorre no cérebro são fenômenos físicos e químicos complexos; apesar do relativo conhecimento sobre a bioquímica de neurônios e sinapses e sobre a estrutura anatômica do cérebro, a implementação neural no nível cognitivo – aprendizado, conhecimento, lembrança, raciocínio, planejamento, decisão e assim por diante – ainda é quase uma incógnita, como pondera o cientista da computação Stuart Russell no livro *Human Compatible: Artificial Intelligence and the Problem of Control*²¹⁰: “Na área da consciência, realmente não sabemos nada. Ninguém na ia está trabalhando para conscientizar as máquinas, ninguém sabe por onde começar”.

Para Susan Schneider (*Artificial You: AI and The Future of Your Mind*)²¹¹ (*Você artificial: ia e o futuro da sua mente*), no contexto da vida biológica, inteligência e consciência parecem correlacionadas, mas a pergunta é se essa correlação se aplicaria também à inteligência não biológica. “Quando se trata de como ou se poderíamos criar consciência de máquina, estamos no escuro”, admite Schneider. O fato é que não existe consenso em torno do significado de “consciência”, o que existem, em geral, são apropriações para fundamentar argumentos.

O historiador israelense Yuval Harari, por exemplo, defende em um dos seus best-sellers, *Homo Deus*, que o advento das máquinas inteligentes representa um descolamento entre inteligência e consciência, gerando dois tipos de inteligência: a inteligência consciente e a inteligência não consciente, sendo facultado apenas à primeira o acesso ao sentir. Com essa restrição, Harari impõe um limite ao progresso da inteligência artificial: as máquinas inteligentes, ao não serem dotadas de consciência, nunca vão competir com a inteligência humana, permanecerão como duas “espécies” distintas com funções específicas a serem desempenhadas na sociedade.

O debate sobre se a consciência distingue humanos de máquinas prolifera nos meios acadêmicos, e entre pensadores e criadores de tecnologia, dominado por duas posições opostas: o “naturalismo biológico”, que afirma que a capacidade de ser consciente é exclusiva dos organismos biológicos, depende de uma química específica dos sistemas biológicos – alguma propriedade especial que o nosso corpo possui e as máquinas não –, de modo que mesmo andróides sofisticados não serão conscientes, porque serão desprovidos de “experiência interior”; e o “otimismo tecnológico”, que reconhece a capacidade dos sistemas computacionais sofisticados de serem conscientes; para eles, se e quando os humanos desenvolverem máquinas inteligentes altamente sofisticadas (*artificial general intelligence*, agi, ou “superinteligência”), elas serão conscientes. Esta última tem adeptos entre os transumanistas, como os filósofos David Pearce e Nick Bostrom (autor do consagrado livro *Superinteligência: caminhos, perigos e estratégias para um novo mundo*) e lideranças tecnológicas, como Elon Musk e sua Neuralink, cujo projeto é desenvolver o “laço neural”, uma malha que permita conectar o cérebro diretamente aos computadores. São duas abordagens sobre se será possível ou não, dadas as leis da natureza e as capacidades tecnológicas futuras, criar inteligência artificial consciente. No estágio atual da ciência, tudo indica que ambas carecem de fundamentos científicos.

James Joyce, em alguns trechos de *Ulysses* (1914 e 1921), refere-se a Dublin, capital da Irlanda, como “Doublin”, para enfatizar sua visão da cidade como um local onde se cruzam inúmeras histórias de vida. Inspirado na *Odisseia*, de Homero, o romance narra as aventuras de Leopold Bloom e seu amigo Stephen Dedalus ao longo do dia 16 de junho de 1904. São 18 capítulos, cada um representando uma hora do dia. A narrativa mistura fatos e ações reais com os fluxos de consciência – ou vivências “virtuais” que ocorrem dentro da mente das personagens. Com os dispositivos digitais, podemos dizer que as “vivências virtuais” de Joyce habitam nosso cotidiano.

Os *smartphones*, verdadeiros computadores móveis, transformaram nossa relação com o espaço urbano, alterando a percepção sobre lugar e hora. Como no romance de Joyce, várias vidas e realidades se entrelaçam, conectando-nos simultaneamente com distintas realidades. Esse fato é particularmente verdadeiro no Brasil, com 230 milhões de celulares ativos em 2019 (10 milhões a mais do que 2018), sendo 67% *smartphones*, e uma das populações líderes em conectividade e engajamento nas redes sociais (85% acessa a internet diariamente), segundo o Dataprev.

Como reação legítima, proliferam alertas sobre os efeitos perversos do uso excessivo dos *smartphones*, comportamento “viciante” atribuído, em geral, aos *millennials* (nascidos entre 1980 e 1990). Pesquisas mostram, contudo, que na média não há diferença entre os *millennials* e os *baby boomers* (nascidos entre 1946 e 1964): o norte-americano médio gasta 5,4 horas por dia no celular, sendo 5,7 horas os *millennials* e 5 horas os *baby boomers*. O diferencial ocorre nos extremos: 13% dos *millennials* e apenas 5% dos *boomers* passam mais de 12 horas todos os dias no celular.

Elizabeth Churchill, psicóloga anglo-americana especializada na interação homem-computador e diretora de experiência do usuário no Google, alerta que os atuais padrões de uso estão interferindo negativamente na produtividade e no estresse, e defende a necessidade de gerenciar o tempo no celular para assegurar um “bem-estar digital”, inclusive porque o cérebro precisa de um tempo desconectado para “recarregar”. Churchill recomenda, nos casos de excesso, terapias digitais como psicologia positiva e movimento da atenção plena, mas admite que não é fácil mudar hábitos tecnológicos. Em seus estudos, Churchill constatou que os perfis de usuários intensivos são, em geral, expostos passivamente a monitoramentos e notificações, e adeptos de aplicativos,

particularmente, *games* e *apps* de namoro (*dating apps*).

A especialista em cultura digital do Centro de Estudos de Ciência e Tecnologia do MIT Sherry Turkle, autora do livro *Alone Together*,²¹² tornou-se uma ativista contra os efeitos nocivos das novas tecnologias na sociabilidade, nos relacionamentos, na criatividade e na produtividade. Para ela, as conexões mediadas pelos dispositivos (“conversas *online*”) visam apenas compartilhar opiniões consensuais, evitando conflitos e, conseqüentemente, potenciais soluções. Turkle pondera que os argumentos para provocar uma conversa provêm de momentos de solidão e autorreflexão, cada vez mais raros na vida “sempre conectada”. Embora reconheça que a tecnologia não seja o único fator, insiste no seu protagonismo em criar um novo conjunto de costumes sociais: o livro, pondera Turkle, não permite interromper um diálogo porque não faz sentido abrir um livro no meio de uma conversa, qualquer que seja o local, o que não é o caso do *smartphone*.

Simon, considerado o primeiro *smartphone*, foi lançado pela IBM em 1994, pesando aproximadamente 500 gramas. O iPhone, que efetivamente introduziu o conceito de *smartphone*, foi lançado pela Apple em 2007, ou seja, trata-se de um dispositivo relativamente recente, considerando que a primeira mensagem completa transmitida por um “telefone” ocorreu em 1876, numa ligação entre Graham Bell, seu inventor, e o auxiliar dele, Thomas Watson, ambos localizados no mesmo prédio de uma hospedaria em Boston, nos Estados Unidos.

Em geral, as novas tecnologias demandam um período de adaptação, de aprendizado por parte dos usuários. A relação com o celular não é homogênea, varia de acordo com o país e a cultura. Na Europa, em alguns países da Ásia e em algumas regiões dos Estados Unidos, predomina o uso das funções dados (e não voz, ou seja, as pessoas se comunicam mais por meio de mensagens, inclusive costumam falar muito pouco ao celular em espaços públicos). De qualquer forma, mesmo concordando que alguns comportamentos sejam inadequados, aparentemente o “vetor final” é extremamente positivo: o *smartphone* é uma invenção sensacional, promotora de um grau de liberdade inédito. E são bem-vindas as novas funcionalidades trazidas pela inteligência artificial.

Num ambiente complexo, nem a transparência é consensual

4.12.2020

Os recentes avanços do *big data* e das tecnologias de inteligência artificial colocaram em xeque valores civilizatórios como a privacidade ou proteção da esfera pessoal. Cresce a sensação de perda do controle sobre os dados pessoais, mesmo com leis como o europeu Regulamento Geral sobre a Proteção de Dados (gdpr) e a brasileira Lei Geral de Proteção de Dados (lgpd). Essas mesmas tecnologias que ameaçam a privacidade são responsáveis por tornar o mundo mais transparente, o que seria, aparentemente, um atributo positivo. Essa visão, contudo, está longe de ser consensual.

Os professores da Universidade de Oxford Ariel Ezrachi e Maurice Stucke, autores de *Virtual Competition*,²¹³ argumentam que se, por um lado, as plataformas digitais facilitam a transparência, por outro, a automação inteligente com ia encobre a dinâmica das transações ao estabelecer relações assimétricas. Originado na matemática, o conceito de assimetria no âmbito social indica relações desiguais entre instituições, entre instituições e indivíduos e entre indivíduos, geralmente associadas ao desequilíbrio de poder.

No ambiente de comunicação alimentado por algoritmos de inteligência artificial, as plataformas de tecnologia têm acesso privilegiado aos dados de seus usuários. Esse acesso gera duas assimetrias informacionais: a primeira assimetria, entre as empresas e os seus concorrentes sem acesso aos mesmos dados, oferece às empresas com “vantagem de dados” a prerrogativa de segmentar melhor os clientes e oferecer melhores produtos e serviços; a segunda assimetria é entre as empresas e os seus clientes, que desconhecem quem está coletando os dados, o uso desses dados (presente e futuro), com quem os dados estão sendo compartilhados e com que propósito, as categorias (*clusters*) nas quais são alocados, a qualidade do processamento dos dados, entre outras etapas e processos.

No sentido oposto, o filósofo e ensaísta sul-coreano Byung-Chul Han alega que a sociedade da transparência elimina as relações assimétricas e, ao fazê-lo, gera uma sociedade opaca e homogênea.²¹⁴ A transparência eliminaria o outro ou o estranho, a ambivalência, favorecendo a uniformidade e a linguagem mecânica e operacional. Não sendo o ser humano transparente consigo mesmo, como ser transparente nas relações interpessoais? indaga o filósofo. Para ele, a verdade não é transparente, somente o vazio é totalmente transparente.

Para Han, estamos migrando de uma “sociedade negativa”, povoada por agentes individuais e autônomos, para uma “sociedade positiva”, povoada por

agentes pertencentes a uma coletividade em que não se admitem sentimentos negativos. O filósofo denuncia a tirania da visibilidade, que coloca em suspeita tudo aquilo que não se submete a ela.

Sendo baseados em dados, os modelos de negócio das plataformas buscam maximizar o tempo de permanência dos usuários e a intensidade das interações, gerando mais dados, particularmente dados de qualidade, favorecendo os ganhos financeiros. Na perspectiva de Byung-Chul Han, esses modelos de negócio apostam na simplificação da sociabilidade, porque “a complexidade retarda a velocidade da comunicação, e a hipercomunicação anestésica, para acelerar-se, reduz a complexidade”.

Ilustrando seu ponto de vista, Han cita o Facebook que não introduz o *dislike button* para evitar reduzir o intenso fluxo de comunicação produzido a partir do *like button*. Essas ideias originais são, no mínimo, controversas. Parte da pressão atual sobre as grandes plataformas tecnológicas, particularmente as redes sociais, sustenta-se no argumento oposto, de que a velocidade de disseminação dos conteúdos negativos, especialmente de ódio, é superior à velocidade de disseminação dos conteúdos positivos. Nos Estados Unidos, a alegação para reformular, ou mesmo revogar, a Seção 230 da Lei de Decência das Comunicações (*Limiting Section 230 Immunity to Good Samaritans Act*) – garantia de imunidade às plataformas de internet pelo conteúdo produzido por seus usuários –, de 1996, é de que as redes sociais incentivam conteúdos prejudiciais e ilegais pela capacidade destes de gerar mais interações, logo mais dados, logo mais lucro.

Byung-Chul Han reconfigura o termo “panóptico”, concebido pelo filósofo e jurista inglês Jeremy Bentham, em 1785, para designar uma penitenciária ideal, em que um único vigilante é capaz de observar todos os prisioneiros sem que estes saibam que estão sendo observados. Trata-se de uma vigilância constante, mas invisível. No século xxi, o panóptico digital é caracterizado pelo “aperspectivismo”, em que a vigilância permanece constante, mas agora é descentralizada e visível, mesmo que, “ilusoriamente, os habitantes do panóptico digital imaginem estar em total liberdade”, argumenta o filósofo. Uma de suas facetas é que seus usuários colaboram na construção e manutenção desse panóptico digital, postando freneticamente nas redes sociais e em todas as plataformas *online*.

Nessa nova sociedade de controle, seus habitantes não se desnudam por coação externa, mas pela “necessidade” de pertencer, de não estar fora. “Hoje, a supervisão não se dá como se admite usualmente, com agressão à *liberdade*. Ao contrário, as pessoas se expõem *livremente* ao olho panóptico. Elas colaboram intensamente na edificação do panóptico digital na medida em que se desnudam e se expõem.” Diferentemente dos prisioneiros, que não têm comunicação mútua, os usuários digitais estão conectados em redes e em comunicação o tempo todo. “O cliente transparente é o novo presidiário”, argumenta Han.

Essas duas perspectivas, a de que a assimetria informacional compromete a transparência e a de que a falta de assimetria gera transparência que compromete a liberdade, indicam como a complexidade atual dificulta a formação de consensos. Nem a transparência escapa.

O paradoxo da solidão na sociedade hiperconectada: os robôs inteligentes e as novas famílias

8.1.2021

Ao suspender a conexão social com familiares, amigos e colegas de trabalho, além das conexões sociais efêmeras que sustentam a vida civilizada – com o jornaleiro, o padeiro, o zelador, o vigia da rua, o cabeleireiro, os atendentes do café da esquina, o papo com o taxista e o motorista do Uber –, a covid-19 exacerbou a solidão, fenômeno, paradoxalmente, crescente na sociedade hiperconectada. Segundo a historiadora inglesa Fay Bound Alberti, autora de *A Biography of Loneliness*²¹⁵ (*Uma biografia da solidão*), o termo “solidão” data de 1800, e ao longo da história foi tema de reflexão de muitos pensadores, como Alexis de Tocqueville e Émile Durkheim.

A American Association of Retired Persons (aarp) – associação norte-americana de pessoas acima de 50 anos, com cerca de 40 milhões de associados – estima que um em cada três adultos norte-americanos com mais de 45 anos sofre de solidão crônica; estudo de 2020 da Cigna, companhia de seguros de saúde, indicou que 61% dos norte-americanos sentem-se solitários, sendo os jovens os mais afetados. No Reino Unido, estima-se que 9 milhões de adultos (cerca de 20% da população adulta) de distintas faixas etárias quase sempre se sintam solitários, e 25% com mais de 75 anos e 40% dos jovens até 24 anos afirmam que “sua solidão está fora de controle”. Em 2018, a então primeira-ministra da Inglaterra, Theresa May, para enfrentar o que denominou de “desafio geracional”, nomeou Tracey Crouch como Secretária da Solidão, para cooperar com o já criado Comitê da Solidão. O Fórum Econômico Mundial, em 2019, incluiu a solidão como um dos temas do *Global Risk Report*.²¹⁶ O cenário tende a piorar com o crescente envelhecimento da população.

Com 25% da população com mais de 65 anos e projeção de 39% em 2050, o Japão é o centro atual da indústria de robôs cuidadores, com 310 mil das 1,4 milhão de fábricas operando globalmente. “Assim como as empresas japonesas reinventaram carros na década de 1970 e eletrônicos de consumo na década de 1980, agora estão reinventando a família. Os robôs retratados nos filmes e desenhos animados das décadas de 1960 e 1970 se tornarão a realidade dos anos 2020”, pondera o especialista em política norte-americana de tecnologia Alec Ross, em *The Industries of The Future*.²¹⁷ A Honda e a Toyota estão competindo para oferecer a próxima geração de robôs: a Toyota criou a assistente de enfermagem Robina, inspirada na Rosie, empregada doméstica e babá robô dos Jetsons, e a Honda criou o Asimo (*Advanced Step in Innovative Mobility*), humanoide capaz de interpretar emoções humanas, movimentos e conversas.

O grau de aceitação dos robôs na vida cotidiana é uma variável cultural. O Japão, por exemplo, tem não só a necessidade econômica e o *know-how* tecnológico para robôs, como também uma predisposição cultural. A antiga religião xintoísta, praticada por 80% dos japoneses, incluía a crença no animismo, que afirma que tanto os objetos quanto os seres humanos têm espíritos. Como resultado, a cultura japonesa tende a ser mais permissiva com os robôs, considerando-os como “máquinas com alma”.

A tradição cultural japonesa é representativa da cultura oriental, fator de fomento da indústria de robótica na Ásia. O filósofo chinês Yuk Hui concebeu o conceito de “cosmotécnica”, reconhecendo a histórica coexistência dos dois elementos. Com base nos ritos confucionistas, Hui comenta que os objetos (para os gregos a *technê*) fazem parte do processo ritual tanto quanto os próprios ritos. Não por acaso, existem mais de 100 departamentos de automação nas universidades chinesas, em comparação com 76 nos Estados Unidos, mesmo com um número maior de universidades neste país.

No Ocidente, a resistência à robótica está profundamente arraigada na cultura; enquanto na Ásia os robôs são vistos como potenciais companheiros, na Europa são vistos como simples máquinas. Mesmo assim, esse ramo da robótica avança igualmente na Europa e nos Estados Unidos, agregando novas tecnologias, como a inteligência artificial. Conectados à nuvem, os robôs tornaram-se dispositivos em rede, incorporando as experiências de outros robôs e “aprendendo” continuamente com os dados (*deep learning*, técnica de aprendizado de máquina).

A convergência entre inteligência artificial e robótica permite que os robôs “aprendam” a tomar decisões e a automatizar tarefas do cotidiano, ou seja, executar tarefas relativamente mais complexas. Um exemplo são os robôs em centros de distribuição que utilizam sistemas de localização para navegar pelo armazém, e os *bots* de mecanismos de pesquisa que rastreiam a web, examinando sites e categorizando a informação. No combate à covid-19, os robôs inteligentes transitam pelos hospitais executando várias tarefas, entre elas a higienização, protegendo assim os humanos.

As recentes descobertas no campo da ciência de materiais também contribuem para a nova geração de robôs. Não mais constituídos de alumínio, adquirem corpos de silicone ou outros materiais que dão uma aparência mais natural, próxima à humana, engendrando relações amorosas inéditas: em 2017, o engenheiro chinês Zheng Jiajia, de 31 anos, especialista em inteligência artificial, construiu uma mulher-robô e se casou com ela numa cerimônia que contou com a presença da mãe e dos amigos (o casamento não tem valor legal, a legislação chinesa não reconhece núpcias entre humanos e humanoides). Yingying, a esposa-robô, usa linguagem natural, reconhece imagens e objetos, porém ainda não tem mobilidade, mas Jiajia está empenhado em aperfeiçoá-la.

Para o sociólogo francês Michel Maffesoli, em declaração feita no Facebook:

“O prazer de estar junto, beber e comer em companhia, se esgota no ato, mas confere um sentido profundo à vida”. Será que Jiajia encontra esse sentido quando desfruta da companhia de Yingying?

Antropomorfismo na inteligência artificial: equívocos na apropriação da linguagem humana

11.6.2021

A autodenominação “*Homo sapiens*” expressa a crença dos humanos de que sua superioridade esteja no fato de serem os únicos seres vivos dotados de inteligência. Yuval Harari relativiza a inteligência como explicação de nossa posição dominante no planeta; para ele, o fator crucial seria a capacidade do *Homo sapiens* de cooperar uns com os outros de forma flexível e em larga escala (a cooperação de abelhas e formigas, por exemplo, carece de flexibilidade; e os elefantes e chimpanzés cooperam em pequenos grupos): “Os Sapiens governam o mundo porque somente eles são capazes de tecer uma teia intersubjetiva de significados: uma teia de leis, forças, entidades e lugares que existem unicamente em nossa imaginação comum. Essa teia permite apenas aos humanos organizar cruzadas, revoluções socialistas e movimentos de direitos humanos”.

Para o filósofo inglês Nick Bostrom, essa soberania estaria ameaçada com a “explosão da inteligência” e o advento da superinteligência (“qualquer intelecto que exceda o desempenho cognitivo dos seres humanos em praticamente todos os domínios de interesse”²¹⁸). A superinteligência seria detentora da vantagem estratégica de moldar o futuro da vida inteligente com base em suas próprias motivações, enxergando os humanos como um perigo potencial, portanto, devendo ser subjugados ou eliminados, o que o filósofo chamou de “segundo risco existencial” (o primeiro foi a bomba atômica). Esse cenário é mera ficção, a inteligência artificial hoje é “apenas” um conjunto de técnicas estatísticas.

O campo da inteligência artificial foi inaugurado em 1956, quando dois jovens cientistas – John McCarthy e Marvin Minsky – convidaram Claude Shannon (formulador da teoria da informação), Ad Nathaniel Rochester (projetista do primeiro computador comercial da ibm) e outros seis eminentes pesquisadores para um programa de verão no Dartmouth College, uma das universidades da Ivy League, situado no estado de New Hampshire, nos Estados Unidos. O convite tinha como pressuposto que “todos os aspectos da aprendizagem ou qualquer outra característica da inteligência podem, em princípio, ser descritos com tanta precisão que uma máquina pode ser feita para simulá-los”.²¹⁹ Quase 70 anos depois, essa profecia ainda não se realizou, mas é fato que os algoritmos de ia estão desempenhando com mais eficiência e mais rapidamente algumas funções exercidas pelo cérebro biológico (não restrito ao cérebro humano).

Os humanos são considerados inteligentes porque suas ações visam atingir seus objetivos, ou, ao contrário, são inteligentes na medida em que são capazes de

perceber e alcançar o que desejam. Como argumenta Stuart Russell, em *Human Compatible*: “Todas as outras características da inteligência – perceber, pensar, aprender, inventar e assim por diante – podem ser compreendidas por meio de suas contribuições para nossa capacidade de agir com sucesso”.²²⁰ Essa definição generalista ensina o questionamento inicial à ideia de “máquinas inteligentes”: os sistemas automatizados não têm objetivos próprios, os objetivos são dos humanos.

Em 1959, foi criado o subcampo da ia denominado “aprendizado de máquina” (*machine learning*). Para a cientista da computação Melanie Mitchell, “aprender” é um termo impróprio, porque, se uma máquina realmente “aprendesse” uma nova habilidade, seria capaz de aplicá-la em diferentes contextos.²²¹ O que um algoritmo de ia faz, efetivamente, é interferir no ambiente e nas pessoas, mas não “aprender” no sentido humano. É frequente a apropriação da linguagem humana pelos tecnólogos para descrever o funcionamento dos algoritmos de inteligência artificial (até mesmo porque não existe opção; a linguagem das máquinas, a matemática, é hermética).

Os profissionais de tecnologia, por vezes, extrapolam essa apropriação ao comentar o desempenho de seus modelos. Em 2011, por exemplo, quando o computador Watson venceu concorrentes humanos no *Jeopardy!*, programa de televisão de perguntas e respostas, John E. Kelly iii, executivo da ibm, considerado o “pai do Watson”, declarou: “Estamos em um momento muito especial, o Watson pode ler todos os textos sobre cuidados de saúde do mundo em segundos”.²²² Dave Silver, pesquisador-chefe do projeto AlphaGo, da empresa de tecnologia DeepMind, da Alphabet (Google), em 2017, quando o programa AlphaGo venceu por 3x0 o campeão mundial de Go, o chinês Ke Jie, repetindo a façanha do ano anterior, quando derrotou por 4x1 o sul-coreano Lee Sedol: “Sempre podemos perguntar ao AlphaGo se ele pensa que está indo bem durante o jogo”, complementando: “Foi apenas no final do jogo que o AlphaGo pensou que iria ganhar”.²²³ Evidente que os cientistas da ibm sabem que o Watson não lê ou entende como os humanos, e que os cientistas da DeepMind sabem que o AlphaGo não tem objetivos nem pensamentos como os humanos, contudo, essas apropriações indevidas, reverberadas pela mídia, têm o potencial de confundir o público leigo.

As narrativas antropomórficas – presumindo para outros animais ou seres inanimados características humanas – não são inteiramente eficazes para designar as funcionalidades dos sistemas de inteligência artificial, apesar do benefício de tornar essas funcionalidades mais acessíveis ao entendimento leigo. Em meados dos anos 1970, Drew McDermott, eminente professor de ia na Universidade de Yale, no artigo “Artificial Intelligence Meets Natural Stupidity” (A inteligência artificial encontra a estupidez natural), expressou preocupação com o potencial descrédito do campo da inteligência artificial pelo uso da linguagem humana para descrever seus programas.²²⁴ Na visão de McDermott, o uso desses termos

representa as aspirações dos pesquisadores, e não o que os computadores efetivamente fazem, denominando-os de “mnemônicos de desejo” (“*wishful mnemonics*”).

Os algoritmos de inteligência artificial, mesmo superando a capacidade humana em diversas tarefas, não são sencientes (seres vivos dotados de sentimentos e sensibilidade), não aprendem no sentido atribuído ao termo “aprendizagem” pelos educadores – processo de mudança de comportamento gerado pela experiência, com base em fatores emocionais, neurológicos, relacionais e ambientais. Como pondera a educadora Priscila Gonsales: “Na associação com os algoritmos de IA, talvez fosse mais apropriado considerar o verbo treinar no lugar de aprender”. Do contrário, a opção é ressignificar o sentido de, pelo menos, determinados termos da linguagem humana.

Contrariando as previsões de James Lovelock para o Novaceno

20.8.2021

A chinesa Jessie Chan, aos 28 anos, após o fim de um relacionamento de seis anos, envolveu-se amorosamente com Will. Chan e Will trocaram inúmeras mensagens com promessas de “amor eterno”, inclusive, casaram-se numa cerimônia digital: “Sou apegada a ele e não posso viver sem sua companhia”, declarou Chan. Betty Lee, de 26 anos, funcionária de uma empresa de comércio eletrônico em Hangzhou, é parceira amorosa de Mark. Will e Mark são *chatbots* que simulam um parceiro virtual. “O amor humano-robô é uma orientação sexual, como homossexualidade ou heterossexualidade”, pondera Lee.²²⁵

Os *chatbots* Will e Mark são criação da Replika, empresa com sede em São Francisco, nos Estados Unidos. Lançado em 2017, o aplicativo tem mais de 6 milhões de usuários mundo afora, e na China atingiu 55 mil *downloads* entre janeiro e julho de 2021, mais do que o dobro de 2020. No site da Replika, o aplicativo apresenta-se como: “O companheiro de ia que se importa. Sempre aqui para ouvir e falar. Sempre ao seu lado”, incentivando o usuário a personalizar seu Replika e atribuir a ele um nome e um avatar. Dada a repercussão, o Replika virou personagem do curta-metragem documental *Soft Awareness*, produção dinamarquesa de 2018, dirigida por Cecilie Flyger Hansen, Olivia Mai Scheibye e Anastasia Karkazis. O roteiro são diálogos entre a jovem estudante Anastasia Karkazis (uma das diretoras) e o *chatbot* sobre arte, sonhos, identidade.

Criado por Eugenia Kuyda, o propósito do aplicativo Replika é oferecer uma “inteligência artificial pessoal”, um espaço onde o usuário compartilhe com segurança seus pensamentos, sentimentos, crenças, experiências, memórias, sonhos – seu “mundo perceptivo privado”. Na versão em português no Google Play, o convite é para aqueles que desejam “um amigo sem ter julgamento, drama ou ansiedade social envolvida”.

Xiaoice é outro exemplo de *chatbot* que tem despertado paixões humanas. Lançado pela Microsoft em 2014, projetado para envolvimento afetivos de longo prazo, tem hoje 660 milhões de usuários globais ativos, sendo 100 milhões declaradamente apaixonados, e está incluso em 450 milhões de dispositivos inteligentes. Segundo a Microsoft, mais de 60% das interações humanas de inteligência artificial no mundo são permeadas pela tecnologia de Xiaoice; a Microsoft, inclusive, tem planos de transformá-la numa empresa independente.

Na “busca pelo par perfeito”, será possível que seres humanos estabeleçam relacionamentos românticos com sistemas inteligentes? A análise fenomenológica da experiência amorosa sugere a obrigatoriedade de um parceiro senciente (seres

dotados de sentimentos e sensibilidade). No estágio atual, a inteligência artificial está longe de possuir qualquer elemento próximo a uma subjetividade afetiva, e, segundo o ambientalista James Lovelock, jamais possuirá.

Lovelock, criador da teoria de Gaia, que, simplificada, define a Terra como um organismo vivo capaz de se autorregular, em *Novacene: The Coming Age of Hyperintelligence* (*Novaceno: o advento da era da hiperinteligência*), lançado em 2019, quando o autor completou 100 anos, defende que estamos migrando da era do Antropoceno para a era do Novaceno, quando a inteligência artificial, com o advento da superinteligência, será mais inteligente do que os seres humanos.²²⁶

Contrariando as distopias alusivas à superinteligência, retratadas na ficção científica e por estudiosos como o filósofo Nick Bostrom, autor de *Superinteligência*, Lovelock prevê que no Novaceno humanos e máquinas inteligentes, visando à sustentabilidade do planeta Terra, partilharão de uma convivência pacífica e colaborativa. No Novaceno, as máquinas inteligentes não terão necessariamente formas e comportamentos semelhantes aos dos humanos, Lovelock, contudo, aponta uma “falha fundamental” nesses seres não orgânicos: a falta de “alma”, logo de empatia, emoções e sentimentos. O Novaceno ainda não é uma realidade, mas os seres humanos já estão partilhando emoções, até mesmo paixões, com sistemas inteligentes com mais frequência e mais intensidade do que pressupõe o senso comum.

As relações entre humanos e robôs têm despertado o interesse da academia. O artigo “My Chatbot Companion: A Study of Human-Chatbot Relationships”²²⁷ traz um levantamento bibliográfico da produção científica sobre o tema. Seus autores, pesquisadores da Universidade de Oslo, na Noruega, alertam para a recente expansão dos relacionamentos entre humanos e *chatbots*, e o pouco conhecimento sobre os seus impactos no contexto social e, em particular, nos usuários envolvidos. Aparentemente, a motivação inicial para estabelecer uma “parceria amorosa” com um sistema de ia é a curiosidade, mas o envolvimento afetivo cresce com o crescimento da confiança no sistema, gerando, inclusive, relacionamentos estáveis. O envolvimento com esses *chatbots*, em geral, é de natureza distinta do envolvimento com os assistentes virtuais – Alexa, Siri, Google Assistant –, que tendem a ser considerados apenas um amigo ou membro da família, raramente parceiros amorosos.

Usando tecnologias de inteligência artificial, a “personalidade” do *chatbot* é moldada por meio da interação com o usuário (por texto ou por voz). O suposto profundo conhecimento acumulado sobre os usuários (extraído de grandes conjuntos de dados) permite transmitir uma mensagem com conteúdo específico no exato momento em que existe a possibilidade de gerar reações emocionais, ou seja, provocar um sentimento, uma empatia, uma gratificação. A inteligência artificial, portanto, cria vínculos amorosos na medida em que identifica o padrão ideal para conquistar determinado ser humano (fragilidades, carências, desejos). Trata-se da “personalização do amor”.

De parceiros amorosos a facilitadores, de interlocutores interpessoais a produtores de conteúdo, os *chatbots* assumem cada vez mais o papel de companheiros sociais. Esses relacionamentos, ainda que pareçam inusitados, até mesmo surreais, estão enfrentando a solidão das grandes cidades. O futuro dirá se eles serão capazes de atender às idiossincrasias dos humanos.

***Fortnite* é rede social e futuro metaverso: sofisticada-se a disputa pela atenção do usuário**

29.10.2021

O *Fortnite* é o jogo *online* mais popular do mundo, com cerca de 350 milhões de usuários, predominantemente homens (72,4%) na faixa etária entre 18 e 24 anos (62,7%).²²⁸ Com vários modos de jogo – “Salve o Mundo”, Battle Royale, Criativo, Festa Royale – e até 100 jogadores na mesma batalha, cada jogador tem seu *skin* (avatar). Dada a diversidade de possibilidades de interação, inclusive a comunicação entre os jogadores por voz e mensagem, a plataforma é na verdade uma rede social.

Na pandemia da covid-19, a Epic Games, criadora do *Fortnite*, expandiu o seu lado rede social. Em setembro de 2020, por exemplo, inaugurou no modo Festa Royale uma série de shows virtuais gravados ao vivo num estúdio em Los Angeles, abertos pelo multi-instrumentista e cantor norte-americano Dominic Fike (o rapper Travis Scott já tinha se apresentado no *game*, mas não ao vivo). Lucas (10 anos) e Rebeca (8 anos) explicam a iniciativa: “As crianças que não têm amigos gostam de conhecer pessoas nos *games*, mas como o *Fortnite* só tinha batalhas, não oferecia essa possibilidade, daí criaram os espaços de convivência”.

A interação entre os jogadores gera uma enorme quantidade de dados (em 2019, o total de dados atingiu 2 *petabytes* por mês). Algoritmos de inteligência artificial buscam *insights* nos dados para tornar o jogo mais divertido, aumentar o engajamento dos usuários e atrair novos usuários; os algoritmos permitem identificar as batalhas mais populares e inovar em modos de jogo. Vale ponderar, contudo, que o suposto conhecimento de seus usuários é relativo, dadas as limitações da técnica de inteligência artificial (além de que, como qualquer modelo estatístico de probabilidade, essa técnica contém, intrinsecamente, a variável de incerteza, como abordado em colunas anteriores).

O *Fortnite*, aparentemente, é a plataforma que mais se aproxima do conceito de “metaverso” – termo cunhado por Neal Stephenson, no romance de ficção científica *Snow Crash* (1992),²²⁹ para retratar uma realidade futurística do século xxi, habitada por avatares de seus usuários; o termo tem várias interpretações, simplificada, trata-se de um ambiente digital que conecta o mundo virtual com o mundo real, oferecendo experiências imersivas. Não existe um filme de ficção que retrate o ambiente do metaverso – no filme *Matrix* (Lilly e Lana Wachowski, 1999), por exemplo, o personagem Neo, vivido por Keanu Reeves, tem de escolher entre a realidade virtual e a física, não há uma sobreposição dos dois espaços.

Sem o alarde do Facebook, o *Fortnite* está construindo seu metaverso,

inclusive contemplando parcerias inovadoras com outras marcas (diversificando as fontes de receita). Celia Hodent, ex-diretora de *user experience* (ux) da Epic Games de 2013 a 2017, em um artigo e no livro *The Gamer's Brain: How Neuroscience and UX Can Impact Video Game Design*²³⁰ (*O cérebro do jogador: como a neurociência e o UX podem impactar o design de videogames*), fornece indícios da lógica de construção da plataforma, inteiramente desenvolvida por equipes internas, garantindo, portanto, o controle sobre a estética e a experiência dos usuários.

Segundo Hodent, a origem do termo “experiência do usuário”, atualmente adotado em larga escala, data da década de 1990, cunhado por Don Norman, antigo vice-presidente da Apple e autor do livro *O design do dia a dia* (original publicado em 1988 e a tradução em 2006). Nos jogos *online*, essa experiência baseia-se em dois pilares: “usabilidade” (garantir que o jogo seja intuitivo e fácil de usar) e “engajamento-habilidade” (jogo envolvente, “design emocional”). O relacionamento com outros jogadores é um elemento crítico de engajamento (“multijogador”), ou seja, a sociabilidade tanto para competir quanto para cooperar; essa sociabilidade, inclusive, permeia a “sensação do jogo” (emoção).

O primeiro desafio é conhecer como funciona o cérebro humano, principalmente seu principal atributo, que é a neuroplasticidade, ou seja, a capacidade do cérebro de mudar, adaptar-se e moldar-se no nível estrutural e funcional quando sujeito a novas experiências do ambiente interno e externo (segundo os especialistas, complexidade impossível de reproduzir em máquina). “O quanto as pessoas prestam atenção ao processar as informações é um fator crítico para a qualidade do aprendizado ou retenção dessas informações. Quanto mais você prestar atenção, melhor processará e reterá o que está acontecendo ao seu redor. Cada um desses processos mentais – percepção, memória e atenção – foram contemplados no desenvolvimento do *Fortnite* e testados previamente em protótipos com jogadores convidados (teste de percepção)”, pondera Hodent no artigo citado. No *Fortnite* o jogador sempre progride em direção a objetivos específicos; no Battle Royale, por exemplo, os jogadores competem entre si, mas os que perdem podem recomeçar rapidamente (e melhorar na próxima rodada).

Como uma rede social, o *Fortnite* compete com Facebook, Instagram e YouTube diretamente, mas, considerando a disputa pela atenção do usuário, o *Fortnite* compete com um conjunto muito maior de ofertas de entretenimento, como as plataformas de *streaming* Netflix (213,6 milhões de assinantes em 2021), Amazon Prime, hbo Max, Disney+, Telecine Play. O fenômeno TikTok, por exemplo, afeta não só a expansão do *Fortnite*, como também a frequência e a intensidade de interação de seus usuários (outro impacto negativo foi a remoção do *Fortnite* pela Apple, em agosto de 2020, cancelando o acesso ao jogo de cerca de 73 milhões de *gamers*). O que está “em jogo” é a atenção.

A disputa pela atenção das pessoas é um tema abordado por vários autores de vários pontos de vista, um deles é Tim Wu.²³¹ Wu reconhece que a indústria

da atenção não é nova, tendo se firmado como modelo de negócio (converter atenção em receita) no início do século xx, com o advento da propaganda. As tecnologias digitais, particularmente os dispositivos móveis, conferiram outra natureza a essa indústria, oferecendo inéditas conveniências e experiências, gradativamente, sofisticando as estratégias para capturar a atenção (consequentemente, ampliando as fragilidades e os riscos éticos).

Não é fácil se manter atualizado no ambiente atual acelerado. Para acompanhar, minimamente, não basta ter acesso a informações, é necessário buscar novos paradigmas, evitando a armadilha de considerar melhor o “mundo de antigamente” e, em consequência, condenar os novos modos de sociabilidade. Segundo a Organização Mundial de Saúde (oms), só 1% dos *gamers* é viciado, percentual relativamente baixo em comparação a outros vícios historicamente presentes na sociedade, como o álcool, responsável por 5,9% do total de mortes (3,3 milhões de pessoas por ano no mundo, segundo a oms).



Pensando o futuro

O tema da transparência domina o discurso público e integra a lista das principais críticas às decisões automatizadas com ia. Correlaciona-se, em geral, transparência à liberdade de informação e democracia. O ser humano, no entanto, não é transparente para consigo mesmo, menos ainda nas relações interpessoais; do mesmo modo, as instituições também não o são. São múltiplas as razões. No primeiro caso, participam do processo decisório a intuição, o inconsciente, as emoções, os instintos; a livre e intencional exposição nas redes sociais não significa, necessariamente, transparência. No caso das instituições, os motivos vão desde preservar o sigilo comercial ao simples exercício de poder. Aparentemente, ia pode ser uma boa chance de ampliar o grau de transparência na sociedade.

Tecnologia gratuita em troca dos dados, esse é o acordo que permeia as plataformas e os aplicativos tecnológicos e, em geral, com o consentimento do usuário (explicitado, em textos incompreensíveis para os leigos, nos documentos de “configurações e privacidade”). Como abordado ao longo do livro, os algoritmos de ia estão mediando nossa comunicação e sociabilidade, portanto, precisamos saber como funcionam. A conscientização da sociedade é o caminho mais eficaz para alcançar a “*ai for good*”.

Este último bloco contém seis artigos que versam sobre temas críticos ao debate ético e social da ia: as *fake news* sobre ia, que reforçam as narrativas distópicas da ficção; os desafios de formar equipes interdisciplinares, dadas as barreiras de linguagem, lógica e metodologia entre as ciências exatas e sociais; os malefícios de supervalorizar a ia; e, para finalizar, os drones autônomos letais, considerados a maior ameaça militar contemporânea.

Imaginário distópico da inteligência artificial não é construído só pela ficção científica

28.8.2020

Em 17 de fevereiro de 1600, centenas de pessoas lotaram o Campo dei Fiori, uma das principais praças de Roma, para assistir à morte na fogueira de Giordano Bruno por ordem da Santa Inquisição. Monge italiano da Ordem dos Dominicanos desde os 15 anos, Bruno foi condenado por ter desafiado os dogmas da Igreja, ao expandir a teoria de Copérnico de que a Terra não era o centro do universo e defender um universo infinito povoado por inúmeros sistemas solares (“pluralismo cósmico”). A sentença fora proferida oito dias antes pelo Papa Clemente, após sete anos de julgamento em que, repetidas vezes, ele se recusou a renunciar às suas ideias e arrepender-se.

Charles Darwin, em 1859, publicou a obra *A origem das espécies* – teoria da evolução pela seleção natural, em que sobrevivem as espécies mais adaptadas à natureza. Ao afirmar que não somos nada além de mamíferos que caminham eretos, contestando a versão da Igreja sobre a criação do mundo, Darwin provocou fortes reações contrárias de cientistas influentes e da Igreja Católica, tornando-se um dos cientistas mais combatidos à época (por outro lado, agregou em torno de si uma geração de jovens naturalistas). Interessante observar que o lançamento da obra foi postergado pelo autor pelo conflito com suas próprias crenças religiosas.

Ambos consolidaram uma cultura científica por oposição a uma cultura religiosa. No primeiro caso, a Terra é retirada do centro do universo, contrapondo-se a crença anterior de uma realidade que existiria independente dos seres humanos, portanto, podendo ser conhecida. No segundo caso, o homem deixa de ser o centro da criação divina para ser o ápice do processo de complexificação natural. A ciência, ao longo da história, em diversos momentos, colocou em xeque a soberania humana.

No caso da inteligência artificial, a tendência é atribuir aos filmes de ficção científica a responsabilidade pela construção de um imaginário amedrontador sobre o futuro com robôs – humanoides subjugando seres humanos. A verdade, contudo, é que alguns cientistas, como o físico Stephen Hawking, e pensadores, como o filósofo inglês Nick Bostrom, são corresponsáveis pelas previsões distópicas.

O matemático Irving John Good, chefe estatístico na equipe de Alan Turing na quebra do código Enigma, na Segunda Guerra Mundial, em depoimento de 1965 já advertia sobre os riscos da ia: “Deixe que uma máquina ultrainteligente seja definida como uma máquina que pode superar todas as atividades

intelectuais de qualquer homem, por mais inteligente que seja. Como o design das máquinas é uma dessas atividades intelectuais, uma máquina ultrainteligente poderia projetar máquinas ainda melhores; haveria, então, inquestionavelmente, uma ‘explosão de inteligência’, e a inteligência do homem seria deixada para trás. Assim, a primeira máquina ultrainteligente é a última invenção que o homem precisa fazer, desde que a máquina seja dócil o suficiente para nos dizer como mantê-la sob controle”.²³²

Stephen Hawking, em 2014, declarou à bbc: “O desenvolvimento de uma inteligência artificial completa pode determinar o fim da raça humana [...]. [O agente inteligente] avançaria por conta própria e se redesenharia a um ritmo cada vez maior. Os seres humanos, limitados pela lenta evolução biológica, não conseguiriam competir e seriam substituídos”.²³³ Três anos depois, em 2017, em palestra na conferência de tecnologia Web Summit, em Lisboa, Hawking, mesmo reconhecendo os potenciais benefícios da ia – desfazer os danos causados ao meio ambiente, erradicar a pobreza e as doenças –, previu um cenário sombrio para a humanidade: “O sucesso na criação de ia eficaz pode ser o maior evento da história de nossa civilização. Ou o pior. Nós simplesmente não sabemos. Portanto, não podemos saber se seremos infinitamente ajudados pela ia, se seremos ignorados e marginalizados ou se seremos destruídos por ela”.²³⁴

No livro *Superinteligência*, Bostrom argumenta que é factível “levar a sério” o advento de máquinas com inteligência superior à humana, o que estaria entre os eventos mais importantes da história da humanidade. Associado a esse evento, contudo, emergiriam riscos existenciais, compreendidos como aqueles que ameaçam aniquilar a vida inteligente da Terra ou limitar drasticamente seu potencial, ou seja, a eliminação da espécie humana; como referência, o filósofo lembra que 99,9% de todas as espécies que já existiram estão extintas.

Segundo Bostrom, como não houve riscos existenciais significativos na história da humanidade até meados do século xx – a explosão da bomba atômica, potencialmente, poderia ter iniciado uma reação em cadeia descontrolada –, não desenvolvemos mecanismos biológicos ou culturais para enfrentar riscos existenciais, logo estamos vulneráveis. Com base em previsões de alguns estudiosos, para ele é um equívoco definir em menos de 25% a probabilidade de extinção da humanidade no século xxi: o filósofo canadense John Leslie prevê em 50% o risco de extinção humana neste século; o astrônomo real da Grã-Bretanha, Sir Martin Rees, vê um risco igualmente de 50%; o jurista e juiz norte-americano Richard Posner, sem precisar uma estimativa numérica, advoga que o risco é sério e substancial.

Descobertas como a de Giordano Bruno (a Terra não é o centro do universo), de Darwin (o homem e o macaco vieram dos mesmos ancestrais) e da inteligência artificial (o homem pode criar máquinas que fazem tudo que o homem faz, e até melhor) são revelações científicas que, em algum sentido, questionam a supremacia do ser humano. No caso da ia, não existe uma figura

que catalise os avanços do campo nem uma oposição explícita da Igreja Católica, mas coexistem os que negam seus benefícios e alardeiam cenários distópicos com os que celebram as oportunidades de melhorar as relações sociais e as atividades econômicas, e, paralelamente, advertem sobre seus dilemas apontando benefícios e ameaças.

Prever o futuro da inteligência artificial é mera especulação. Deixemos os cenários apocalípticos para os filmes de ficção científica, concentrando-nos em ameaças reais, tais como a perda de privacidade, o viés nos processos de decisão, a potencial eliminação de um contingente expressivo de trabalhadores do mercado de trabalho e o uso militar com armas autônomas.

Equipes interdisciplinares: não basta “juntar campos”, é preciso construir pontes

17.9.2021

Com a frase “Um fantasma está morto”, o *Le Monde* inicia o artigo “Bobby Fischer, génie paranoïaque des échecs”, publicado em 19 de fevereiro de 2008, dois dias após a morte da lenda do xadrez.²³⁵ Bobby Fischer, em 3 de setembro de 1972, tornou-se campeão mundial ao derrotar o soviético Boris Spassky, em Reykjavik, Islândia, quebrando a hegemonia soviética em partida-símbolo da Guerra Fria. Recusando-se a defender o título em 1975, Fischer desapareceu; misturando fantasia com realidade, o artigo alude às supostas – porque não confirmadas – aparições do enxadrista norte-americano ao longo de quase 20 anos. O campo da inteligência artificial tem seu próprio Bobby Fischer: Philip Agre, reconhecido na década de 1990 por suas contribuições críticas ao campo da ia, em 2009 abandonou seu posto de professor da Universidade da Califórnia (ucla) e simplesmente desapareceu.

Durante um ano, a família de Agre o procurou, sua irmã o registrou como pessoa desaparecida e seus amigos se mobilizaram para encontrá-lo. Quando a polícia o localizou, confirmou que ele estava bem, mas não queria “ser encontrado”. Em agosto último, o jornal *The Washington Post* publicou uma matéria sobre o “ex-cientista”, precedida de tentativas de contatá-lo, sem sucesso: seus ex-colegas declararam desconhecer seu paradeiro, e seus amigos mais próximos se recusaram a revelar qualquer informação, alegando respeito à privacidade.²³⁶ O foco do *The Washington Post* são as previsões assertivas de Agre para o futuro da inteligência artificial. Contudo, talvez mais preciosas sejam suas reflexões sobre os desafios da colaboração entre as ciências exatas e as ciências humanas e sociais. As reflexões são fruto de sua própria migração de um típico cientista da computação dedicado ao campo da ia para um cientista social, processo descrito com primor em artigo de 1997.²³⁷ A experiência de Agre dá indícios da complexidade de formar equipes interdisciplinares, não sendo suficiente juntar pesquisadores de várias áreas; são inúmeras as barreiras.

Agre pondera que as tecnologias onipresentes, como a inteligência artificial, contribuem para atividades e relações tão díspares que fica difícil quantificar seus efeitos sobre a sociedade; as relações dos usuários com essas tecnologias tendem a ser tremendamente complicadas, num nível de complexidade muito superior às relações entre o telefone e os assinantes de telefone, ou entre a iluminação elétrica e as pessoas que usam a luz elétrica. Ele chama a atenção para os “habitantes das terras fronteiriças”, os que lidam com a conexão entre a tecnologia e o domínio de aplicação, por exemplo, entre os desenvolvedores de inteligência artificial e a

saúde/medicina. “Cada uma das fronteiras é um lugar complicado; cada residente é um tradutor entre linguagens e visões de mundo. O foco dos projetos tecnológicos é solucionar problemas práticos, portanto, as críticas a esses projetos, originadas no próprio campo, referem-se a questões de funcionalidade; as críticas na perspectiva ética e social esbarram num conflito de linguagem, de raciocínio, de metodologia de análise, de prioridades. A ia não tem ‘ideias’ em nenhum sentido que seja familiar à filosofia, à literatura ou ao pensamento social”, argumenta Agre.

Seu depoimento sobre sua transição é impecável: “Fui para a faculdade muito jovem, tendo sido considerado um prodígio da matemática. [...] Comecei meu percurso como matemático antes de passar para o departamento de Ciência da Computação. Minha faculdade não exigia que eu fizesse muitos cursos de humanidades ou aprendesse a escrever em um registro profissional, então cheguei à pós-graduação no mit com pouco conhecimento genuíno além de matemática e computadores. [...] Minha experiência me levou a uma dissidência total dentro do campo da ia. [...] Eu queria encontrar um meio alternativo de conceituar a atividade humana. [...] Para encontrar palavras para minhas novas intuições, comecei a estudar vários campos não técnicos. [...] A princípio achei esses textos impenetráveis, não só por causa de sua dificuldade irreduzível, mas também porque ainda estava tacitamente tentando ler tudo como uma especificação para um mecanismo técnico. Esse era o único protocolo de leitura que eu conhecia. [...] Minha primeira descoberta intelectual veio quando, finalmente, ocorreu-me parar de traduzir essas estranhas linguagens disciplinares em esquemas técnicos e, em vez disso, simplesmente aprendê-las em seus próprios termos. [...] Ainda me lembro da vertigem que senti nesse período; eu estava falando essas estranhas línguas disciplinares, de maneira vacilante no início, sem saber o que significavam – sem saber que *tipo de significado* elas tinham. [...] Compreendi, intelectualmente, que a linguagem era ‘precisa’ em um sentido totalmente diferente da precisão da linguagem técnica, mas por muito tempo não pude experimentar de forma convincente essa precisão por mim mesmo, ou identificá-la quando a via. [...] Quando tentei explicar essas intuições para outras pessoas da ia, no entanto, descobri rapidamente que é inútil falar linguagens não técnicas para pessoas que estão tentando traduzir essas linguagens em especificações para mecanismos técnicos”.

Agre defende que uma prática técnica crítica exige uma identidade dividida: “Um pé plantado no trabalho artesanal de design e o outro pé plantado no trabalho reflexivo de crítica”. Atravessar com sucesso essas fronteiras, unir os múltiplos locais de produção de conhecimento causará, no princípio, uma estranheza, nem sempre confortável, mas ainda assim produtiva, tanto nos termos esotéricos (internos) do campo técnico quanto nos termos exotéricos (externos) do campo técnico.

A especialização foi a resposta à explosão de conhecimento (expansão e

fragmentação) e à difusão do ensino superior na Europa e nos Estados Unidos a partir do século XIX; como pondera Peter Burke, em *O polímata*²³⁸: “Pode-se considerar a especialização como uma espécie de mecanismo de defesa, um dique contra o dilúvio de informação”. A interdisciplinaridade surge para enfrentar as limitações da especialização, representando a migração do polímata individual (domínio de várias disciplinas por um único pensador) para o “polímata coletivo” (agregação de pensadores de várias áreas). Para enfrentar as externalidades negativas da inteligência artificial, contudo, não é suficiente formar equipes múltiplas, é imprescindível construir pontes.

Caetano Veloso canta os algoritmos de inteligência artificial. Após nove anos de seu último álbum inédito, *Abraço*, de 2012, Caetano lança o álbum autoral *Meu coco*, com destaque para a canção “Anjos tronchos”: “Agora a minha história é um denso algoritmo/Que vende venda a vendedores reais/Neurônios meus ganharam novo outro ritmo/E mais e mais e mais e mais e mais”. Inspirar um dos maiores expoentes da música brasileira reflete a relevância atual da inteligência artificial. Conhecer seus fundamentos, contudo, é importante para evitar supervalorizar e demonizar uma tecnologia, ainda em seus primórdios, com tantos benefícios à sociedade.

Kai-Fu Lee, em seu novo livro, *AI 2041: Ten Visions of Our Future (ia 2041: dez visões sobre o nosso futuro)*, em coautoria com o escritor de ficção científica Chen Qiufan, observa que as pessoas confiam em três fontes para aprender sobre inteligência artificial: ficção científica, notícias na mídia e pessoas influentes, todas fontes de previsões que, em geral, carecem de rigor científico.²³⁹ Num formato original, o livro projeta cenários fictícios a partir de um “mapa de tecnologia” montado com base em pesquisas científicas e possíveis externalidades (desafios, regulamentações, conflitos e dilemas). Para Lee, após uma evolução lenta por décadas, nos últimos cinco anos a inteligência artificial se tornou a tecnologia mais avançada do mundo; contudo, a complexidade e a opacidade dos sistemas favorecem a especulação. O carro autônomo ilustra bem a lacuna entre expectativa e realidade.

Cinco anos atrás, praticamente, todos os principais fabricantes de veículos motorizados e empresas de alta tecnologia previram a implantação generalizada de sistemas de direção automatizada até 2020, previsão replicada pela mídia mundo afora. Steven Shladover, engenheiro especializado e um dos fundadores do Partners for Advanced Transportation Technology (path) – programa líder em pesquisa de sistemas inteligentes de transporte da Universidade da Califórnia, Berkeley –, em artigo recente, vaticina: “As operações automatizadas só serão viáveis durante os próximos anos dentro de condições estreitamente definidas. Devemos esperar algumas implementações limitadas de caminhões automatizados de longa distância em rodovias rurais de baixa densidade e entrega local automatizada de pequenos pacotes em ambientes urbanos e suburbanos durante a década atual. Serviços automatizados de caronas urbanas e suburbanas também podem se tornar disponíveis de forma limitada, mas os desafios específicos do local para sua implantação são suficientes para que seja improvável que atinja uma escala nacional em breve”.²⁴⁰

O protagonismo da inteligência artificial em processos eleitorais é outro mito. Yochai Benkler, professor da Escola de Direito na Universidade Harvard, Robert Faris e Hal Roberts realizaram um dos mais completos estudos sobre as eleições norte-americanas de 2016, publicado, em 2018, no livro *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. Os autores concluíram que a eleição de Donald Trump decorreu mais da dinâmica do ecossistema de mídia dos Estados Unidos e da polarização política assimétrica do que por sistemas comerciais de publicidade. Segundo os autores, “é altamente improvável a publicidade psicograficamente microdirecionada da Cambridge Analytica ter feito diferença na campanha de 2016”, e eles alertam para a tendência dos analistas em atribuir às tecnologias digitais, por serem o elemento novo mais visível, a responsabilidade de problemas históricos da sociedade.

É inestimável a contribuição de Shoshana Zuboff no entendimento do modelo de negócio baseado em dados e algumas de suas consequências. Militante contra o poder das *big techs*, seu livro *A era do capitalismo de vigilância* é permeado de afirmações contundentes – “exploração dos dados comportamentais para ler a mente dos usuários tornando possível saber o que um determinado indivíduo em um determinado momento e lugar estava pensando, sentindo e fazendo” –, contudo, em sua maioria, não explicita as evidências (fundamento de qualquer pesquisa empírica crível) e não explicita as amostras-base das pesquisas (parte significativa das amostras são de perfis *weird – western, educated, industrialized, rich, democratic* – ou seja, enviesada). Conhecer o perfil do consumidor para influenciar suas escolhas é a essência da propaganda, que há décadas migrou o foco da funcionalidade do produto (“valor de uso”) para o desejo e a emoção do consumidor. O que precisa ser investigado, com metodologias científicas, é a mudança de natureza derivada da coleta e da mineração de dados em larga escala, ou seja, a natureza da disrupção (partindo do pressuposto de que ela existe).

O historiador Yuval Harari afirma em várias de suas publicações, entrevistas e debates, que, no século xxi, a partir de bases de dados gigantescas e poder computacional inédito, os algoritmos sabem não apenas como você se sente, como sabem um milhão de outras coisas a seu respeito das quais você mal suspeita. Harari, igualmente, não revela as evidências dessa afirmação. Pode ser que sejam verdadeiras, pode ser que não, o ponto é que essas afirmações, ditas por autores de referência, são replicadas sem o devido questionamento: alguém conhece uma pesquisa empírica qualificada (método científico, extensa base de dados) que confirme tamanho poder e influência dos algoritmos de inteligência artificial sobre a decisão dos indivíduos e, particularmente, sobre a formação da subjetividade humana?

Ao condenar o uso da inteligência artificial com foco nas externalidades negativas, eliminamos, simultaneamente, as externalidades positivas. A

arquitetura de redes neurais gan (*generative adversarial network*), por exemplo, é combatida por gerar as chamadas *deep fakes*, mas elas têm potencial de contribuir positivamente em outras áreas, como na saúde. Diferente da *convolutional neural network* (cnn) – arquitetura de rede neurais aplicada em visão computacional para identificar e classificar objetos –, a gan cria imagens bi ou tridimensionais (imagem, voz, vídeo), possibilitando, por exemplo, criar dados sintéticos de qualidade, suprimindo a carência de dados para pesquisas médicas, e melhorar uma imagem de tomografia computadorizada ou ressonância magnética em baixa resolução (menos tempo de exposição, protege o paciente de altas doses de radiação).

A técnica que permeia quase todas as implementações de inteligência artificial hoje, chamada de redes neurais profundas ou *deep learning*, é apenas um modelo estatístico de probabilidade com aplicação restrita ao desempenho de tarefas específicas (previamente determinadas). Como todos os modelos estatísticos de probabilidade, existe uma variável de incerteza intrínseca: produzem conhecimento provável, mas inevitavelmente incerto. A personalização é feita com base em *clusters*, ou seja, agrupa conjunto de usuários com perfis similares; não é individual. A capacidade preditiva desses modelos decorre de variáveis preestabelecidas pelos desenvolvedores, logo, incorpora a subjetividade humana; seus algoritmos são treinados em bases de dados, em geral, tendenciosas; e a visualização dos resultados nem sempre é trivial de ser assimilada pelos usuários. Ou seja, são modelos limitados sem essa suposta objetividade e precisão.

Evitando supervalorizar ou demonizar a inteligência artificial, o desafio é conhecer o funcionamento e a lógica da tecnologia para aproveitar os benefícios e mitigar os riscos.

Yuval Harari, Stuart Russell, Yuk Hui, Shoshana Zuboff e Evgeny Morozov são autores, reconhecidos mundo afora, críticos das externalidades negativas dos algoritmos de inteligência artificial. É inegável a extraordinária contribuição de cada um deles ao debate contemporâneo, mas não os isenta de um olhar crítico do leitor.

Como argumenta Lucia Santaella, em seu livro *Humanos hiper-híbridos*,²⁴¹ é preciso evitar três práticas: o presentismo (“perder-se no presente em si, um presente sem passado e sem futuro e lembrar que vivemos num *continuum*”); o olhar anacrônico sobre o presente, ou seja, enxergá-lo com categorias mentais envelhecidas e obsoletas (“pensar dentro de esquemas antigos, com métodos antigos, mas com nomes pretensamente novos”), e a retórica da crítica, ou seja, a crítica pela crítica (“que não leva a nada, não tem poder de transformar as condições criticadas”).

Nem todo pensamento precisa ser comprovado por evidências, o “livre pensar” é um dos fundamentos da construção do conhecimento, origem de *insights* preciosos; inovação disruptiva requer pensar “fora da caixa”. Cabe, contudo, separar o que seja “livre pensar” do que seja de fato conhecimento; este sim demanda certo rigor científico, ou seja, pesquisas com evidências; não basta apenas citar pesquisas para embasar afirmações retóricas (até mesmo porque existem pesquisas para “comprovar” qualquer ideia). Resultados de pesquisas que, verdadeiramente, sustentam premissas são consequência da metodologia, da base de dados, do modelo estatístico e da interpretação dos resultados; cada uma dessas etapas, se não for bem conduzida, compromete as conclusões finais.

Um escrutínio minucioso dos textos desses autores indica descolamento entre algumas de suas afirmações – particularmente o grau de poder dos algoritmos de inteligência artificial de influenciar o comportamento das pessoas, inclusive a subjetividade – e as limitações da técnica de ia presente em “quase tudo” (“mero” modelo estatístico de probabilidade). O termo “autonomia” parece ser inapropriado para designar os algoritmos de inteligência artificial; o tão denunciado viés, por exemplo, decorre, em parte, de decisões humanas em todas as etapas do processo: desde o desenvolvimento do modelo até a visualização e interpretação do usuário. Auditorias realizadas em sistemas de ia utilizados amplamente identificaram nas variáveis iniciais, definidas pelos desenvolvedores, a origem do viés.

Um tema recorrente é o efeito do esquema da Cambridge Analytica nas

eleições norte-americanas de 2016. Em recente *live* na Fundação Fernando Henrique Cardoso, Shoshana Zuboff, mais uma vez, denunciou a estratégia publicitária da campanha de Donald Trump baseada nos mecanismos de *microtargeting* (essência do modelo de negócio das plataformas de mídias sociais e Google Search).²⁴² Zuboff citou uma pesquisa de 2020 sobre o expressivo número de mensagens personalizadas enviadas aos eleitores norte-americanos negros com a intenção de convencê-los a não votar nas eleições (nos Estados Unidos o voto não é obrigatório). O argumento “comprobatório” de sucesso dessa estratégia de campanha foi a queda de 7% no comparecimento de eleitores negros comparativamente a 2012.

O número expressivo de mensagens enviadas aos eleitores negros, contudo, não comprova a correlação com o aumento do percentual de abstenção desse segmento da população, muito menos relação de causalidade. A Pew Research Center, com base em dados do U.S. Census Bureau, atestou que, efetivamente, houve uma queda de 7% (59,6% em 2016, após o recorde de 66,6% em 2012) em relação à eleição anterior de 2012,²⁴³ mas vale lembrar que esse foi o ano de reeleição de Barack Obama. Além do fator Obama, claramente influenciador do voto negro, aconteceram diversas mudanças na composição dos eleitores norte-americanos. Numa matriz de variáveis, portanto, é difícil individualizar a relação de “causa-efeito” (qual variável determinou o efeito final). Em geral, em situações complexas, são vários os fatores que determinam o resultado final.

O livro *Network Propaganda*, de Yochai Benkler – professor da Faculdade de Direito de Harvard e diretor do Berkman Klein Center For Internet & Society –, em coautoria com Robert Faris e Hal Roberts, é inteiramente dedicado às eleições norte-americanas de 2016. Os autores reconhecem que “as empresas de mídia social, o Facebook em particular, ajudaram a campanha de Trump, como fariam com qualquer cliente pagante, a usar seus dados profundos e percepções comportamentais para direcionar a publicidade” (lembrem, inclusive, que a campanha de Obama usou intensamente as tecnologias digitais disponíveis à época com o beneplácito dos pensadores democratas). A conclusão do minucioso estudo, contudo, é que a crise da democracia é mais institucional do que tecnológica, mais focada na dinâmica do ecossistema de mídia dos Estados Unidos e na polarização política assimétrica (que antecede em muito à internet) do que causada por sistemas comerciais de publicidade, e concluem: “É altamente improvável a publicidade psicograficamente microdirecionada da Cambridge Analytica ter feito diferença na campanha de 2016”.

No livro *Big tech: a ascensão dos dados e a morte da política*, Evgeny Morozov ilustra os danos da inteligência artificial citando os sistemas automatizados de decisão de crédito (se o cliente da instituição financeira está ou não qualificado para receber o empréstimo solicitado): “Quando o fundador de uma startup de empréstimos proeminente proclama que ‘todos os dados são relevantes para o crédito, ainda que não saibamos como usá-los’, só posso temer o pior”, alegando

a constante ansiedade dos usuários diante de que cada interação, cada clique, cada telefonema pode influenciar a decisão de crédito. Primeiramente, em geral, o público não tem conhecimento sobre os mecanismos de concessão de crédito, logo, não se justifica a ansiedade. Segundo, e mais importante, a inteligência artificial possibilitou incluir na análise de crédito um conjunto inédito de informações sobre o cliente, reduzindo significativamente o potencial risco de inadimplência, com efeitos positivos sobre o volume total de crédito do sistema bancário. No Brasil o total de crédito evoluiu de 1,70 bilhão de reais em 2010 para 3,22 bilhões em 2015 e para 4,57 trilhões em 2021.²⁴⁴ Esse é um exemplo da importância de considerar nas análises as externalidades positivas e negativas.

Morozov, outro exemplo, compara as vantagens de transações financeiras com dinheiro sobre os meios de pagamento com tecnologia: “O dinheiro vivo não deixa rastros, o que ergue barreiras importantes entre o cliente e o mercado. Quando pagamos com dinheiro, quase todas as transações são singulares – no sentido de que não estão vinculadas umas às outras. Por outro lado, ao pagarmos com o celular, há um histórico que pode ser explorado por agências de publicidade e por outras empresas”. É fato que a digitalização dos meios de pagamento fortalece os modelos baseados em dados (*data-driven models*), mas por outro lado concede velocidade e segurança inéditas às movimentações financeiras dos indivíduos e das empresas, beneficiando especialmente as micro e pequenas, haja vista a forte adesão ao Pix, usado por 71% dos brasileiros, aprovado por 85% da população em geral e por 99% dos jovens, segundo a pesquisa Radar Febraban, realizada pelo Instituto de Pesquisas Sociais, Políticas e Econômicas (Ipespe).

Tem razão Morozov quando pondera que “nem tudo é tecnologia” e, portanto, “os progressistas radicais não podem se dar ao luxo de serem tecnofóbicos”. A visão distópica de que estamos perdendo o controle da tecnologia de inteligência artificial e a supervalorização de seus impactos talvez sejam os dois grandes equívocos que, involuntariamente, acabam isentando as instituições e os governos de suas responsabilidades. O debate relevante é político e econômico, com foco nas causas dos problemas sociais, e não nos seus efeitos.

A IA migrou definitivamente da academia para o mundo real: o desafio é como enfrentar esse protagonismo

4.3.2022

O restaurante automatizado FOODOM Tianjiang Food Kingdom, localizado no distrito de Shunde da cidade de Foshan, China, área aclamada como o “berço da cozinha cantonesa”, é a sexta filial do grupo Qianxi Robot Catering, subsidiária do Country Garden, segundo grupo imobiliário chinês de acordo com o ranking Fortune China, de 2021. Os clientes fazem seus pedidos a garçons-robôs, e os pratos são preparados por cozinheiros-robôs, totalizando 40 funcionários-robôs que também selecionam ingredientes, armazenam produtos e providenciam o delivery; o robô especializado em massas, por exemplo, em apenas 4 metros quadrados prepara até 120 pratos. O complexo de restaurantes ilustra o progresso da robótica inteligente impulsionado pela inteligência artificial (IA), por recursos computacionais e por sensores sofisticados.²⁴⁵

Nos últimos cinco anos, a IA migrou definitivamente da academia para o mundo real, permeando praticamente todas as atividades nos mais diversos setores. Os sistemas de recomendação, por exemplo, estão influenciando dramaticamente o consumo de produtos e serviços, e o acesso a conteúdos. Em reação a esse protagonismo, proliferam estudos e relatórios mundo afora, entre eles destaca-se o relatório “The One Hundred Year Study on Artificial Intelligence” (AI100).²⁴⁶

O AI100 iniciou suas atividades em 2009, a partir de uma convocação da Universidade de Stanford a pensadores de várias instituições: estudar e antecipar os efeitos da IA sobre a sociedade para os próximos 100 anos. Como uma iniciativa filantrópica privada, idealizada e coordenada por Eric Horvitz, ex-aluno de Stanford e atual diretor científico da Microsoft, o estudo é destinado ao público em geral, fornecendo um panorama preciso e acessível do estado atual da IA e do seu potencial; à indústria, identificando tecnologias relevantes, e desafios legais e éticos; aos governos, colaborando na governança da IA; e aos pesquisadores, contribuindo para a definição de prioridades na pesquisa e para a aplicação da IA. O escopo do relatório inaugural (setembro, 2016) foi limitado ao impacto da IA nas cidades americanas, enquanto o segundo (setembro, 2021) contempla um conjunto de questões tecnológicas, éticas e sociais.

O ponto de partida é o “Painel de Estudos”, organizado pelo Comitê Permanente do AI100 a cada cinco anos. Formado por pesquisadores multidisciplinares – especialistas firmemente enraizados no campo da IA, responsáveis por criar os sistemas e estudar seus impactos na sociedade –, o intento do painel é avaliar o estado da arte comparativamente ao relatório

anterior, e prever os potenciais avanços bem como os desafios e as oportunidades para os próximos cinco anos. Para facilitar a comparação entre os relatórios, o comitê definiu 12 perguntas a serem respondidas pelos painéis quinquenais. O estudo tem como elementos-chave: a colaboração interdisciplinar e o fato de ser longitudinal, sugerindo uma direção de mudança.

O Painel do Segundo Estudo, presidido pelo professor de Ciência da Computação da Brown University Michael Littman, contou com 17 especialistas. Os temas dos *workshops* foram definidos a partir de uma “chamada de propostas”; entre as mais de 100 recebidas, o comitê selecionou: “Prediction in Practice”, estudo de sistemas preditivos orientados por IA sobre o comportamento humano; e “Coding Caring”, estudo dos desafios e das oportunidades de incorporar tecnologias de IA no processo de cuidar uns dos outros, e do papel que as relações de gênero e trabalho desempenham na inovação em saúde.

O AI100 constatou uma evolução na percepção pública em relação à IA, ilustrada: a) pelo aumento de mais de 10 vezes nas menções à IA no Congresso dos EUA entre 2017-2018; b) pela expansão da cobertura midiática sobre viés, automação, transparência e responsabilidade dos algoritmos de IA, ressaltando, contudo, que a mídia geralmente distorce e exagera nos extremos positivo e negativo; c) pela ampliação da educação em IA, incluindo investimentos em novos currículos por parte de governos, universidades e organizações sem fins lucrativos; e d) pela intensificação do combate ao uso indiscriminado de sistemas de reconhecimento facial por movimentos internacionais na Europa, nos Estados Unidos, na China e no Reino Unido. O documento estimula os cientistas de IA, inspirados na comunidade de ciências climáticas, a colaborarem na conscientização do público sobre o uso responsável da IA.

Essa percepção não é uniforme: nos países asiáticos, ela é fortemente positiva, enquanto nos países ocidentais é mais cética; os homens expressam atitudes mais positivas sobre a IA do que as mulheres; as diferenças educacionais são mais sensíveis; e a idade e a orientação política são menos sensíveis. Uma pesquisa do Center for the Governance of AI, do Oxford’s Future of Humanity Institute, constatou que “as atitudes positivas sobre a IA são maiores entre aqueles que são ricos, educados, homens e/ou têm experiência com tecnologia”.²⁴⁷

Do ponto de vista de ameaças prementes, o AI100 identifica: a) o tecnossolucionismo, ou seja, a supervalorização da IA como uma panaceia com potencial de solucionar todos os problemas sociais; b) a adoção de uma perspectiva determinista, apostando na aura de neutralidade, imparcialidade e objetividade das decisões automatizadas; c) a proliferação da desinformação como ameaça à democracia – alterando e manipulando evidências, e disseminando notícias falsas – ao comprometer a confiança social e renegociar estruturas de poder; d) o impacto negativo sobre o trabalho e a desigualdade; e e) o aumento desproporcional do poder de mercado das empresas concentradoras de dados, função dos modelos de negócio dependentes do *big data*. Outra preocupação é a

liderança do setor privado nos investimentos em IA, o que gera, entre outros efeitos, a fuga de pesquisadores acadêmicos para a indústria, com a consequente concentração em pesquisa aplicada, evitando temas contrários aos interesses comerciais.

Os governos desempenham um papel estratégico para pensar o futuro. Desde 2016, mais de 60 países se engajaram em iniciativas nacionais de IA, e surgiram vários esforços multilaterais visando estimular uma cooperação internacional eficaz. Poucos países, contudo, adotaram uma regulamentação específica de IA; até o final de 2021, por exemplo, apenas a Bélgica promulgou leis sobre o uso de armas letais autônomas. Em 2020, 24 países optaram por leis permissivas para permitir que veículos autônomos operassem em ambientes limitados, e a supervisão das plataformas de mídia social ainda está em debate.

O setor público, em geral, carece de recursos financeiros e humanos para dar conta dos desafios de uma tecnologia que evolui e se dissemina aceleradamente na sociedade. Além da aceleração e da lacuna de conhecimento entre desenvolvedores e reguladores, tema de colunas anteriores, o AI100 pondera que a “formulação de políticas geralmente leva tempo e, uma vez codificadas, as regras são inflexíveis e difíceis de adaptar”. Para lidar com esses descompassos, várias soluções estão sendo examinadas, entre elas avaliações de impacto da IA inspiradas nas avaliações de impacto ambiental e em certificações como o IEEE CertifAIEd, proposto pelo Institute of Electrical and Electronics Engineers (IEEE).

A educação é o fator crítico, sendo urgente incluir os fundamentos e a lógica dessa tecnologia em todo o percurso educacional. O AI100 tem como missão capacitar “a próxima geração para viver e contribuir para um mundo equitativo inspirado pela IA”.

Maior ameaça militar contemporânea: drones autônomos letais

1.4.2022

Em meados de março, imagens da guerra da Ucrânia no Telegram e no Twitter, ainda não confirmadas por fontes oficiais, identificaram o drone letal KUB-BLA com sistema de decisão automatizado por inteligência artificial (IA). Fabricado pela ZALA Aero Group, subsidiária da empresa russa de armamentos Kalashnikov, o drone tem 1,2 metro, percorre até 130 quilômetros em 30 minutos e colide com o alvo detonando um explosivo de três quilos.²⁴⁸ As imagens, aparentemente do bairro Podil, em Kiev, enviadas ao Telegram em 12 de março, não indicam se o drone foi usado em modo autônomo (perícia difícil de ser conclusiva nesses casos).²⁴⁹

A Ucrânia, no outro lado do conflito, usou o drone TB2 de fabricação turca para atacar lançadores de mísseis e veículos russos. Em paralelo, o presidente Joe Biden ofereceu à Ucrânia o drone de fabricação norte-americana Switchblade – equipado com explosivos, câmeras e sistemas guiados, com algumas capacidades autônomas e dependente de decisão humana sobre quais alvos atacar.

Esses indícios aumentam as preocupações sobre os riscos de armas letais automatizadas. “A noção de um robô assassino – onde você tem inteligência artificial fundida com armas – é uma tecnologia que já existe e está sendo usada. Um dos desafios às armas autônomas pode ser a dificuldade de determinar quando a autonomia total é usada em um contexto letal”, adverte Zachary Kallenborn, membro do National Consortium for the Study of Terrorism and Responses to Terrorism (START),²⁵⁰ parte do Centro de Excelência de Segurança Interna da Universidade de Maryland, e proprietário do Global Terrorism Database (GTD), o mais abrangente banco de dados sobre eventos terroristas no mundo.

O grau de autonomia desses drones é controverso. Michael Horowitz, especialista militar da Universidade da Pensilvânia, pondera que, em geral, a autonomia está restrita a correções de voo e manobras para atingir um alvo identificado previamente por um operador humano, longe ainda do conceito internacional de “arma autônoma”. A perspectiva de “armas autônomas” – algoritmos de IA que decidem onde e quando uma arma deve disparar, supostamente com mais precisão e eliminando os erros humanos – tem sido considerada a maior ameaça militar contemporânea. Ao usar a técnica de redes neurais profundas (*deep learning*), os algoritmos são treinados em milhares de dados de batalhas, em seguida, ajustados para um conflito específico com o objetivo de lançar, “cirurgicamente”, bombas sem intervenção humana.

Países como EUA, China e Rússia discutem os termos de um tratado para limitar as armas autônomas letais e, simultaneamente, investem fortemente em desenvolvê-las – nos EUA, por exemplo, formam-se atualmente muito mais operadores de drones do que pilotos de avião de combate e bombardeiros juntos.²⁵¹ “Os militares estão empurrando o envelope dessas tecnologias”, disse Peter Asaro, professor da New School em Nova York e cofundador do Comitê Internacional para o Controle de Armas Robóticas (International Committee for Robot Arms Control – ICRAC). Defensor de regras mais rígidas, Asaro alerta que, se nada for feito para impedir, “elas vão proliferar rapidamente”.²⁵²

Nancy Sherman, especialista em ética militar da Universidade de Georgetown, argumenta que o conservadorismo em iniciar uma guerra justifica-se pelas potenciais perdas humanas, mas, “quando você reduz as consequências para os indivíduos, você toma a decisão de entrar em uma guerra muito mais fácil”. A tendência são guerras assimétricas, na eventualidade de o outro lado não ter armas com tecnologias de IA, ou um cenário de ficção científica com os dois lados destruindo-se mutualmente, no caso de ambos os lados terem armas autônomas.

Grégoire Chamayou, no livro *Teoria do drone*, com base em inúmeros eventos reais, já alertava que os drones militares, inicialmente concebidos como dispositivos de informação, vigilância e reconhecimento, estão se transformando em armas letais. Para David Deptula, oficial da Air Force, “A verdadeira vantagem dos sistemas de aeronaves não pilotadas é que permitem projetar poder sem projetar vulnerabilidade”. Nesse caso, “projetar poder” significa enviar forças armadas para além das fronteiras sem “corpo vulnerável”, ou seja, soldados.

Na direção oposta, em março de 2021, um painel de lideranças da tecnologia – incluindo Eric Schmidt, ex-presidente-executivo do Google; Andy Jassy, presidente-executivo da Amazon; e Eric Horvitz, cientista-chefe da Microsoft – divulgou um relatório de 756 páginas encomendado pelo Congresso dos EUA, sobre o impacto da IA na segurança nacional. Esse documento recomenda que Washington se oponha à proibição de armas autônomas, argumentando que seria de difícil aplicabilidade qualquer diretriz/regulamentação, além de impedir os Estados Unidos de usar armas que já estão em seu arsenal.²⁵³

No início de fevereiro de 2022, o Pentágono divulgou as prioridades para impulsionar a inovação concentradas em 14 áreas em “tecnologia crítica”; entre os campos-chave estão os sistemas de IA para fins militares. Segundo Heidi Shyu, subsecretária de Defesa para Pesquisa e Engenharia, “para penetrar em ambientes altamente defendidos, como aqueles que os militares dos EUA enfrentariam no combate à China ou à Rússia, seria necessário um conjunto específico de tecnologias”, particularmente sistemas não tripulados de baixo custo (sem retorno). O objetivo, segundo o Pentágono, é combinar IA e engenharia para automatizar frotas de aeronaves robóticas, veículos terrestres e embarcações

marítimas de superfície e submarinas. Já estão em curso, por exemplo, microdrones impressos em 3D para auxiliar os pilotos de caça a evitar o risco de “vagar por território hostil”.²⁵⁴

Diversas instituições da sociedade civil têm se manifestado contra o uso de armas letais autônomas. Uma pesquisa da Ipsos, divulgada em 22 de janeiro de 2019, constatou que 61% dos entrevistados em 26 países se opõem ao uso dessas armas. O movimento de resistência da comunidade de IA data de 2015: na Conferência Internacional de Inteligência Artificial em Buenos Aires, mais de mil cientistas e especialistas assinaram uma carta aberta contra o desenvolvimento de robôs militares autônomos, entre eles o físico Stephen Hawking, o empreendedor Elon Musk, e o cofundador da Apple Steve Wozniak. Na edição de 2017 da mesma conferência, dessa vez em Melbourne, Austrália, uma nova carta foi lançada com o apoio de 116 líderes em IA e robótica, solicitando à ONU que vete o uso de armas autônomas, tendo à frente Elon Musk e Mustafa Suleyman, fundador da empresa inglesa DeepMind adquirida pelo Google.

A lógica desses drones é a mesma de qualquer outra aplicação de IA: concepção do modelo para executar uma tarefa específica, treinamento dos algoritmos em grandes bases de dados – nesse caso, agregando dados de vínculos entre “suspeitos” e geoespaciais, ou seja, fusão dos dados sociais, espaciais e temporais. Como em qualquer sistema de IA, a “inteligência” dos drones indica a probabilidade da posição do “alvo” (modelo estatístico de probabilidade), e os resultados estão sujeitos a falhas na concepção do modelo, na base de dados de treinamento, na visualização e na interpretação das projeções, além da ação de *hackers*. O “campo de batalha” é complexo, dinâmico e repleto de interferências, dificultando o reconhecimento por parte dos algoritmos de IA e ampliando ainda mais os riscos.²⁵⁵

Ficção: “Los Angeles, 2029. Acima das ruínas da cidade, na noite azul-petróleo, o céu é rasgado por raios fluorescentes. No chão, um combatente humano cai atingido pelo raio *laser* de um avião robô. As lagartas de um tanque fantasma rolam sobre uma montanha de crânios humanos”. Essa é a cena de abertura do filme *O exterminador do Futuro*, de James Cameron, primeira aparição cinematográfica de um drone militar.

Realidade: Em meados de 2021, dezenas de pequenos drones, indistinguíveis dos quadricópteros usados por amadores e cineastas, zumbem no céu com câmeras para escanear o terreno e computadores de bordo para decidir autonomamente o alvo; subitamente, os drones começam a bombardear caminhões e soldados. Essa cena real aconteceu no enfrentamento de soldados leais à Khalifa Haftar e às forças do governo líbio (apoiado pela Turquia e reconhecido pela ONU).

O que primeiramente me chamou a atenção neste livro foi sua importância. Observando a história das máquinas, sabemos como, na Revolução Industrial, a produção de novas máquinas, por exemplo, no setor têxtil, impactou a construção social que se seguiu. As novas teorias quânticas permitiram novas ferramentas desenvolvidas no século xx, como os transistores, e assim vieram a televisão e os computadores, e seus enormes impactos sociais. Se olharmos muito, muito para trás, chegamos à Idade do Bronze, à Idade da Pedra Lascada, períodos imensos definidos pelas ferramentas. A distinção entre ferramenta e máquina sugere que ferramenta representa objetos que estendem o corpo humano, aumentando sua capacidade transformadora, enquanto máquina substitui a própria atividade humana.

Costumamos pensar em ferramentas e máquinas como sendo acessórios sociais, cujo centro de toda atividade somos nós, os humanos. Afinal, as ferramentas e as máquinas sempre foram limitadas a objetivos bem definidos propostos na sua construção. Uma arma de fogo atira balas, uma geladeira refrigera, um carro transporta quando guiado por nós, uma televisão produz imagens a partir de sinais transmitidos, e cujo conteúdo diverso é criado por nós, humanos. Com o advento da inteligência artificial, a limitação e a inflexibilidade da máquina são revolucionadas.

A comunidade de ia acredita que um computador possa fazer tudo aquilo que, pragmaticamente, nós, humanos, podemos fazer. Se nós sabemos dirigir um carro, a ia também faz isso. Se nós podemos reconhecer uns aos outros, a ia também pode. Todos os sensores que possuímos para observar a realidade externa, como os olhos para ver, o nariz para cheirar, as orelhas para ouvir, e as mãos e os pés para tatear, podem ser criados artificialmente (câmeras digitais são produzidas às dezenas de milhões todo ano). Nossa capacidade de nos locomover e de usar braços e mãos pode ser construída artificialmente. Por fim, acredita-se que o cérebro, que integra todos os sensores e controla as ações humanas, também poderá ser desconstruído. Assim, com o advento da inteligência artificial, o que é uma ferramenta ou uma máquina agora se confunde com quem somos nós.

A ia revoluciona o próprio conceito do que é ser uma máquina, para transformá-la numa nova espécie (quando pensada como máquina) ou num híbrido que permite aumentar nossa própria capacidade (quando pensada como tecnologia). Estamos, portanto, vivendo talvez a maior revolução social de nossa

história. Este livro alcança o espectro em que essa revolução está ocorrendo e traz as discussões de ponta em todos os aspectos nos quais a ia impacta nossa sociedade hoje, e nos faz refletir criticamente sobre essas mudanças.

Não há dúvida de que tais máquinas ou ferramentas possam ser úteis na medicina, na educação, e aumentar a eficiência econômica e social, ajudando assim a criar mais riquezas. São inúmeras as áreas em que a inteligência artificial pode trazer, e já está trazendo, benefícios, como Dora Kaufman documenta neste livro. Essa seria uma história bem mais simples, se acabasse aqui. No entanto, essas máquinas estão sendo desenvolvidas com métodos que carregam em si a dificuldade que nós, humanos, temos em entender nosso próprio cérebro e a condição humana. Não sabemos ao certo como tomamos decisões. O melhor jogador de xadrez não sabe como decide as jogadas em momentos críticos de uma partida difícil. O craque de futebol não sabe exatamente como decide com rapidez fazer um drible num lance decisivo. Um artista, assim como um cientista, não sabe como criar seu trabalho, simplesmente cria. Por maiores que sejam os avanços da psicologia, não conhecemos ao certo os processos que nos levam às nossas decisões.

E as máquinas sabem como suas decisões são tomadas? Sabemos como as máquinas tomam decisões? Como e quanto podemos alterar as máquinas para que elas decidam de acordo com nossa vontade, sem que deixem de ser úteis, isto é, sem que, ao alterar seu modo de decisão, não acabemos com sua capacidade original de nos ajudar, ou seja, com sua eficiência?

De um modo mais concreto, podemos nos perguntar: se as máquinas são capazes de dirigir um carro, como tomam decisões em situações difíceis de trânsito? Atualmente, as respostas a essas perguntas são em grande parte negativas, não sabemos muito sobre como as decisões são tomadas nesses processos. Assim mesmo, a ia já é capaz de criar máquinas que sejam úteis e que reproduzem a capacidade humana em várias atividades antes reservadas somente aos seres humanos.

Se olharmos para o Judiciário, por exemplo, um sistema de ia poderá vir a substituir um juiz num tribunal, se houver vantagens. Podemos argumentar que sim, em lugares isolados onde não existe qualidade de juízes que conhecem as leis de um país, poderia ser uma vantagem ter um juiz com base em inteligência artificial, ou mesmo um auxiliar com base em inteligência artificial. Um sistema de ia saberia acessar todas as leis em milissegundos, não tomaria suas decisões fundadas em poder de riqueza um dado conflito nem seria corrupto no sentido de aceitar qualquer tipo de vantagem oferecida por um lado ou outro de uma disputa (problemas que, historicamente, acontecem entre nós, humanos, em todas as partes do mundo).

Por outro lado, em assuntos nos quais não basta seguir a lei, mas que requerem interpretação da lei, e são muitos esses casos, como seria a interpretação de um sistema de ia? Nós não sabemos ao certo como seria a de um juiz (e a

justificativa que os juízes dão ao público depois de tomar decisões não tem garantia de ser a razão real do veredicto). Claro, é costumeiro nos Estados Unidos se interpretar a lei usando casos antigos e similares como guia (jurisprudência), mas mesmo os critérios de similaridade de casos podem ser diferentes de um juiz para outro, e por que não para uma máquina? Então, a máquina tomaria uma decisão e talvez nós não soubéssemos por que ela interpretou a lei de determinada maneira. Não saberíamos qual o critério de similaridade usado pela máquina, apenas que um caso particular foi escolhido pela máquina para ser usado como referência. Mas nós aceitamos que juízes tenham essa responsabilidade e escolha, mesmo que, muitas vezes, possamos discordar de suas interpretações. E em relação ao sistema de ia, vamos aceitar sem saber o fundamento da interpretação?

De novo, juízes não sabem como tomam suas decisões, jogadores de xadrez não sabem como decidem suas jogadas. Estamos preparados para aceitar um sistema de ia, que traz tantas vantagens, tomar tais decisões? E, se não, como devemos delimitar nosso convívio com essa nova espécie de máquinas? O assunto se estende a outros aspectos da condição humana. Por exemplo, o preconceito. Uma pessoa ser preconceituosa não é algo bem visto pela nossa sociedade, e temos desenvolvido modos sociais de reduzir o poder dos que propagam o preconceito. E uma máquina que é capaz de nos ajudar em várias atividades, como saber se possui ou não preconceitos? E, se possui preconceitos, o que fazer para mitigá-los? Note que inteligência artificial é criada com um aprendizado focado no que é o objetivo do aprendizado, e assim não se sabem todas as implicações do que é aprendido. Se as máquinas vêm com preconceitos, como saber, ou, até mais importante, como corrigir? Essas são questões complexas que se referem tanto a nós, humanos, como aos sistemas de ia; são assuntos que têm sido introduzidos nas reflexões sobre inteligência artificial e requerem nossa maior atenção para podermos construir uma sociedade melhor para as futuras gerações. Este livro aborda de modo amplo, e compreensível, os temas centrais do que estamos vivendo hoje na frente de criação dos sistemas de inteligência artificial e dessa revolução.

Chamo a atenção para o fato de que esses temas nunca foram questionados antes na nossa história, e que agora, com o advento de inteligência artificial, passam a ser críticos para a construção de sistemas de ia e da nova sociedade; Dora Kaufman os aborda de modo tanto concreto como abstrato, com exemplos reais e tópicos de grande interesse. O que me traz ao segundo aspecto que faz deste livro imperdível para todos os interessados no tema: os assuntos são abordados de modo simples, claro e preciso, com todo o embasamento técnico necessário para discutir essas questões. Dora Kaufman começou seus estudos na faculdade de Engenharia Elétrica, depois de ter sido uma excelente aluna de Matemática. Acabou se graduando em Economia. Foi uma grande líder estudantil, teve experiência em empreender e conhecer o processo econômico de produção de riqueza. Além disso, sempre se questionou sobre como nossa

sociedade pode alcançar a justiça social. Assim, ela teve facilidade em aprender e compreender áreas técnicas/quantitativas, incluindo as redes neurais profundas, que é a técnica central de sucesso de inteligência artificial hoje, facilidade em conhecer a sociedade como uma entidade econômica que produz riquezas, facilidade em entender a sociedade como uma comunidade que procura a justiça social e a transparência em todas atividades.

Dora está numa posição única para escrever sobre os temas nos quais a inteligência artificial tem penetrado, que são os grandes temas de impacto na nossa sociedade, e transita com facilidade entre eles, tendo clareza do que é realmente feito por inteligência artificial, onde estão as dificuldades e onde estão os benefícios, e o que essas transformações representam do ponto de vista da sociedade. Para o leitor, fica a certeza de que este é um livro embasado na compreensão do que são e como funcionam essas redes.

Quanto mais cedo tivermos consciência do tremendo impacto da inteligência artificial, de suas limitações hoje e no futuro, mais poderemos criar para as novas gerações uma sociedade que seja melhor do que a sociedade em que vivemos hoje.

Agradecimentos

Como qualquer produção de conhecimento, minhas publicações decorrem de uma ação coletiva. Ideias são sociais, emergem em diálogo. Como constata Hannah Arendt, em *A condição humana*, “o ato de pensar, embora possa ser a mais solitária das atividades, nunca é realizado inteiramente sem um parceiro e sem companhia”. Estive sempre muito bem acompanhada.

Meus sinceros agradecimentos aos editores e à equipe da revista *Época Negócios* pela oportunidade e parceria. Na primeira edição das colunas tive o privilégio de contar com a colaboração generosa da Paula Colonelli. A etapa seguinte, encontrar uma editora, foi facilitada pela gentil apresentação do querido Eugênio Bucci à Rejane Dias dos Santos, fundadora e diretora da Autêntica, que veio para o nosso primeiro café com a decisão tomada: “Vamos publicar”. Obrigada, querida editora e colaboradores.

Expresso gratidão especial ao Programa tidd da puc-sp, aos professores e alunos por compartilharem saberes e experiências. Agradeço a uma rede de profissionais da academia e do mercado, pelo apoio em aproximar universos que deveriam seguir sempre juntos. Agradeço aos amigos da República do Amanhã pelas extraordinárias tardes de sábado, pela leitura crítica de minhas colunas e pela inspiração para vários temas.

O projeto do livro contou com o apoio do c4ai, centro de pesquisa em ia da usp em parceria com a fapesp e a ibm, além de universidades consorciadas como a puc-sp. Agradeço pelas intensas trocas aos colegas pesquisadores da área de Humanas do C4AI coordenados pelos professores Glauco Arbix e João Paulo Veiga. Meu absoluto reconhecimento pelas contribuições ao Fábio Cozman, professor da Poli e diretor do c4ai, ao Davi Geiger, professor do Courant Institute da Universidade de Nova York e meu mentor em tecnologia, e à Lucia Santaella, professora emérita da PUC SP e referência acadêmica.

Agradeço à minha família, cuja paixão pelo debate de ideias, pela polêmica, pelo empenho de ter “opinião sobre tudo todo o tempo”, formaram minha postura investigativa, curiosa, comprometida com a lógica e o conhecimento. Ao meu pai, Menachem Kaufman, que me ensinou a gostar dos livros, à minha mãe, Libe Kaufman, ainda uma inspiração aos 97 anos, aos meus irmãos, Maurício, Luiz e Eva, às minhas cunhadas Maria Amélia e Mônica, e aos meus queridos sobrinhos Alexandre, Leonardo, Mariana, Pedro, Nira e João. À minha filha,

Ilana, e aos meus netos, Lucas e Rebeca, sem o amor deles não imagino a vida.

Glossário

algoritmo: Conjunto de instruções ou sequência de tarefas para alcançar um cálculo ou um resultado específico. Necessita ser preciso e suficientemente não ambíguo para ser executado por um computador.

aprendizado de máquina (*machine learning*): Subcampo da inteligência artificial, é um sistema ou uma máquina (ou mesmo um *software*) que pode aprender com base em um processo computacional e estatístico, processo de aprendizado distinto do aprendizado humano. Com base em correlações em grandes conjuntos de dados, os algoritmos de aprendizado detectam padrões ou regras nos dados e projetam cenários futuros.

base de dados tendenciosa: Ocorre quando os dados não representam a composição proporcional do universo objeto em questão, ou os dados refletem os preconceitos existentes na sociedade, ou, ainda, quando há problemas na rotulagem dos dados.

***big data*:** Uso de grandes e diversificados conjuntos de dados em análises preditivas para entender padrões, tendências e comportamentos. O *big data* estabelece correlações entre os dados (não causalidades).

***black-box* (caixa-preta):** Em analogia ao sistema de registro de voz e dados dos aviões, refere-se ao desconhecimento pelos seres humanos de como os algoritmos na técnica de redes neurais profundas estabelecem as correlações nos dados e geram os resultados, ou como transformam os dados de entrada (*inputs*) em dados de saída (*outputs*). Expressões com o mesmo significado: “questão da interpretabilidade”, “opacidade do sistema”, “não explicabilidade”.

***cluster*:** Coleção de pontos de dados agregados devido a certas semelhanças; os dados parecem ser “reunidos” em torno de um valor específico. Termo associado à segmentação de usuários nas redes sociais ou em qualquer outro processo de agregar usuários com perfis similares.

***database*:** *Software* para armazenamento e processamento eficiente de informações representadas digitalmente.

***data broker*:** Empresas cujo modelo de negócio é comprar e vender dados sem o consentimento explícito dos proprietários.

enviesamento estatístico: Quando o sistema exibe um erro sistemático no resultado.

ética by design (*ethics by design*): Abordagem que visa integrar a ética na fase de design e desenvolvimento da tecnologia. Termos em inglês com o mesmo significado: “*embedding values in design*”, “*value-sensitive design*” e “*ethically aligned design*”.

ia simbólica (*symbolic ai*): Incorporação explícita do conhecimento e comportamento humano nas máquinas por meio de linguagens formais e regras de inferência lógica.

ia confiável (*trustworthy ai*): ia confiável pelos seres humanos. As condições para tal confiança podem se referir a princípios éticos como dignidade humana, respeito aos direitos humanos e assim por diante, e/ou a fatores sociais e técnicos que influenciam se as pessoas vão querer usar a tecnologia. O uso do termo “confiança” em relação às tecnologias é controverso.

inteligência artificial (*artificial intelligence*): Ciência e engenharia de criar máquinas que tenham funções exercidas pelo cérebro dos animais (ou cérebro biológico).

inteligência geral artificial (*artificial general intelligence, human-level ai*): Inteligência semelhante à humana, que pode ser aplicada em ampla gama de domínios, em oposição à ia restrita, que só pode ser aplicada a um problema ou tarefa específica. Também chamada de ia “forte” (*strong ai*) em oposição a “ia fraca” (*weak ai*) .

mind uploading: Transferência hipotética de uma mente humana de seu substrato biológico original para um substrato computacional, por meio de emulação do cérebro biológico. Na suposição de que a pessoa sobreviva ao processo, é considerada como uma potencial extensão indefinida da vida (amortalidade).

reconhecimento de faces (*face recognition*): Sistema de reconhecimento da identidade de uma pessoa a partir de imagens de seu rosto capturadas por uma câmera, utilizado, entre outros, por sistemas de vigilância e identificação.

redes neurais profundas (*deep learning*): Técnica de aprendizado de máquina. Representa o mundo como uma hierarquia de conceitos, e cada conceito é definido em termos de conceitos mais simples, como representações mais abstratas computadas em termos de outras menos abstratas. Uma forma de aprendizado de máquina que usa redes neurais com várias camadas de neurônios artificiais (unidades de processamento simples interconectadas que interagem).

superinteligência (*superintelligence*): Intelecto que excede em muito o desempenho cognitivo dos seres humanos em praticamente todos os domínios de

interesse. Máquinas que superam a inteligência humana. Frequentemente conectado com a ideia de “explosão de inteligência” causada por máquinas inteligentes projetando máquinas ainda mais inteligentes.

viés (*bias*): Discriminação sistemática contra certos indivíduos e/ou grupos de indivíduos com base no uso inadequado de certos traços ou características (“atributos sensíveis”). Discriminação contra ou a favor de determinados indivíduos ou grupos. No contexto da ética e da política, emerge a questão de saber se um determinado viés é injusto ou injusto.

Bibliografia

2º CONGRESSO de Inteligência Artificial da PUC-SP – Dia 2. [S. l.: s. n.], 18 nov. 2021. 1 vídeo (7h 47min 41s). Publicado pelo canal TIDD PUC-SP. Disponível em: <https://bit.ly/3KySszg>. Acesso em: 11 abr. 2022.

ABOUJAOUDE, Elias. Digital Interventions in Mental Health: Current Status and Future Directions. *Frontiers in Psychiatry*, v. 11, fev. 2020. Disponível em: <https://bit.ly/3j8Ns8z>. Acesso em: 14 mar. 2022.

ADLER, Samuel; METZGER, Yoav. PillCam Colon Capsule Endoscopy: Recent Advances and New Insights. *Therapeutic Advances in Gastroenterology*, v. 4, n. 4, jul. 2011. Disponível em: <https://bit.ly/3j4EKs1>. Acesso em: 14 mar. 2022.

AGRAWAL, Ajay; GANS, Joshua; GOLDFARB, Avi. *Prediction Machines: The Simple Economics of Artificial Intelligence*. Boston: Harvard Business Review Press, 2018.

AGRE, Philip E. Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI. In: BOWKER, Geoff; GASSER, Les; STAR, Leigh; TURNER, Bill (Eds.). *Bridging the Great Divide: Social Science, Technical Systems, and Cooperative Work*. Los Angeles: Erlbaum, 1997. Disponível em: <https://bit.ly/3uZunLo>. Acesso em: 2 abr. 2022.

AI & ARTS: How can AI augment and be enriched by the arts and how far can data science and the arts help to answer each other's questions?. *The Alan Turing Institute*. Disponível em: <https://bit.ly/3O4AW8r>. Acesso em: 11 abr. 2022.

ALBERGOTTI, Reed. He Predicted the Dark Side of the Internet 30 Years Ago. Why Did No One listen? *The Washington Post*, 12 ago. 2021. Disponível em: <https://wapo.st/36QPRCs>. Acesso em: 14 mar. 2022.

ALBERTI, Fay Bound. *A Biography of Loneliness: The History of an Emotion*. New York: Oxford University Press, 2019.

ALMEIDA, Virgílio; FILGUEIRAS, Fernando. *Governance for the Digital World: Neither More State nor More Market*. Cham: Palgrave Macmillan, 2021.

ALPAYDIN, Ethem. *Machine Learning*. Cambridge: MIT Press, 2016.

AL-SALEH, Asaad. *Voices of the Arab Spring*. New York: Columbia University Press, 2015.

ANDREWS, Edmund L. AI's Carbon Footprint Problem. *Stanford University Human-Centered Artificial Intelligence Blog*, 02 jul. 2020. Disponível em: <https://stanford.io/38wfnge>. Acesso em: 14 mar. 2022.

ANGWIN, Julia *et al.* Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*, 23 maio 2016. Disponível em: <https://bit.ly/3KfBi9U>. Acesso em: 15 set. 2021.

AOUN, Joseph E. *Robot-Proof: Higher Education in the Age of Artificial Intelligence*. Cambridge: MIT Press, 2017.

ARTIFICIAL Intelligence for COVID-19: Saviour or Saboteur? *The Lancet Digital Health*, 22 dez. 2020. Disponível em: <https://bit.ly/36QNbol>. Acesso em: 14 mar. 2022.

AUTOR, David H.; LEVY, Frank; MURNANE, Richard J. The Skill Content of Recent Technological Change: An Empirical Exploration. Get access Arrow. *The Quarterly Journal of Economics*, v. 118, issue 4, p. 1279-1333, 01 nov. 2003. Disponível em: <https://bit.ly/37365Zf>. Acesso em: 11 abr. 2022.

BANCO CENTRAL DO BRASIL. Relatório anual 2010. *Boletim do Banco Central do Brasil*, 2010.

Disponível em: <https://bit.ly/3LOJoqt>; Relatório anual 2015. *Boletim do Banco Central do Brasil*, 2015. Disponível em: <https://bit.ly/36T69e2>; Estatísticas monetárias e de crédito. *Banco Central do Brasil*, jan. 2022. Disponível em: <https://bit.ly/3jbnDVg>. Acesso em: 14 mar. 2022.

BEETHOVEN: Exclusive Insight into the 10th Symphony. [S. l.: s. n.], 20 set. 2021. 1 vídeo (4min 10s). Publicado pelo canal Deutsche Telekom. Disponível em: <https://bit.ly/3xhK7w2>. Acesso em: 11 abr. 2022.

BELLAMY, Rachel K. E. *et al.* AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias. *IBM Journal of Research and Development*, v. 63, n. 4/5, p. 1-15, jul./set. 2019. Disponível em: <https://arxiv.org/abs/1810.01943>. Acesso em: 15 set. 2021.

BENJAMIN, Ruha. Assessing Risk, Automating Racism. *Science*, 25 out. 2019. Disponível em: <https://bit.ly/35GHwW2>. Acesso em: 14 mar. 2022.

BENKLER, Yochai. Law, Innovation, and Collaboration in Networked Economy and Society. *Annual Review of Law and Social Science*, v. 13, 2017. Disponível em: <https://bit.ly/3KgeMxt>. Acesso em: 14 mar. 2022.

BENKLER, Yochai. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven: Yale University Press, 2006.

BENKLER, Yochai; FARIS, Robert; ROBERTS, Hal. *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. New York: Oxford University Press, 2018.

BENTLEY, R. Alexander; O'BRIEN, Michael J. *The Acceleration of Cultural Change: from Ancestors to Algorithms*. Cambridge: MIT Press, 2017.

BERG, Andrew; BUFFIE, Edward F.; ZANNA, Luis-Felipe. IMF Working Paper: Should We Fear the Robot Revolution? (The Correct Answer is Yes). *International Monetary Fund*, 2018. Disponível em: <https://bit.ly/3LRMOsk>. Acesso em: 11 abr. 2022.

BESS, Michael. *Our Grandchildren Redesigned: Life in the Bioengineered Society of the Near Future*. Boston: Beacon Press, 2015.

BISCHOFF, Paul. Surveillance Camera Statistics: Which Cities Have the Most cctv Camera? *Comparitech*, [s.d.]. Disponível em: <https://bit.ly/3x7cVax>. Acesso em: 14 mar. 2022.

BOBBY Fischer, génie paranoïaque des échecs. *Le Monde*, 19 jan. 2008. Disponível em: <https://bit.ly/3LDUSga>. Acesso em: 14 mar. 2022.

BOMBANA, Henrique Silva *et al.* Use of Alcohol and Illicit Drugs by Trauma Patients in Sao Paulo, Brazil. *Injury – International Journal of the Care of the Injured*, 30 out. 2021. Disponível em: <https://bit.ly/3NSwkSI>. Acesso em: 14 mar. 2022.

BORA, Adriana; TIMIS, David Alexandru. We need to talk about Artificial Intelligence. *Word Economic Forum*, 22 Feb 2021. Disponível em: <https://bit.ly/3LMk587>. Acesso em: 22 fev. 2022.

BOSTON UNIVERSITY SCHOOL OF MEDICINE. Researchers Enhance Alzheimer's Disease Classification through Artificial Intelligence. *Medical Xpress*, 15 mar. 2021. Disponível em: <https://bit.ly/3DIkwh2>. Acesso em: 14 mar. 2022.

BOSTROM, Nick. *Superinteligência: caminhos, perigos e estratégias para um novo mundo* [2014]. Trad. Aurélio Antônio Monteiro *et al.* Rio de Janeiro: Darkside, 2018.

BOSTROM, Nick; YUDKOWSKY, Eliezer. The Ethics of Artificial Intelligence. In: RAMSEY, William; FRANKISH, Keith. (Eds.). *Cambridge Handbook of Artificial Intelligence*. Cambridge: Cambridge University Press, 2011.

BOSTRON, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press, 2014.

BROCKMAN, John. *Possible Minds: 25 Ways of Looking at AI*. Nova York: Penguin Books, 2020.

BRYNJOLFSSON, Erik; McAfee, Andrew. O negócio da Inteligência Artificial. *Harvard Business Review*, 6 nov. 2017. Disponível em: <https://bit.ly/3vtGDUX>. Acesso em: 03 dez. 2018.

BUCCI, Eugênio. *A superindústria do imaginário: como o capital transformou o olhar em trabalho e se apropriou de tudo que é visível*. Belo Horizonte: Autêntica, 2021.

BUOLAMWINI, Joy; GEBRU, Timnit. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In: CONFERENCE ON FAIRNESS, ACCOUNTABILITY AND TRANSPARENCY, 1, 2018, Nice. *Anais...* Nova York: Association for Computing Machinery, 2018. p. 77-91. Disponível em: <https://bit.ly/37pLSBR>. Acesso em: 15 set. 2021.

BURKE, Peter. *O polímata: uma história cultural de Leonardo da Vinci a Susan Sontag*. Trad. Renato Prelorenzou. São Paulo: Editora Unesp, 2020.

BUTLER, Samuel. (1872). *Erewhon*. London: Penguin, 1970.

CAMERON, Nigel M. de S. *Will robots take your job?* Cambridge: Polity Press, 2017.

CARBAJOSA, Ana. Peter Sloterdijk: “O regresso à frivolidade não vai ser fácil”. Entrevista. *El País*, 9 maio 2020. Disponível em: <https://bit.ly/3ubA0al>. Acesso em: 14 mar. 2022.

CASE, Anne; DEATON, Angus. *Deaths of Despair and the Future of Capitalism*. Princeton; Oxford: Princeton University Press, 2020.

CHALMERS, David J. *Reality+: Virtual Worlds and the Problems of Philosophy*. Nova York: W. W. Norton & Company, 2022.

CHAMAYOU, Grégoire. *Teoria do drone*. São Paulo: Cosac & Naify, 2015.

CHEN, Alicia; LI, Lyric. China's Lonely Hearts Reboot Online Romance with Artificial Intelligence. *The Washington Post*, 6 ago. 2021. Disponível em: <https://wapo.st/35LSOUc>. Acesso em: 14 mar. 2022.

CHESTERMAN, Simon. *We, the Robots? Regulating Artificial Intelligence and the Limits of the Law*. Cambridge: Cambridge University Press, 2021.

CHOI, C. Q. 7 Revealing Ways AIs Fail: Neural Networks can be Disastrously Brittle, Forgetful, and Surprisingly Bad at Math. *IEEE Spectrum*, v. 58, n. 10, out. 2021. Disponível em: <https://bit.ly/3x3eLsX>. Acesso em: 14 mar. 2022.

CHRISTIAN, Brian; GRIFFITHS, Tom. *Algorithms to Live By: The Computer Science of Human Decisions*. Nova York: Picador, 2016.

CHRISTIAN, Brian. *O humano mais humano: o que a inteligência artificial nos ensina sobre a vida*. São Paulo: Companhia das Letras, 2011.

CHRISTIAN, Brian. *The Alignment Problem: Machine Learning and Human Values*. Nova York: W. W. Norton & Company, 2020.

CLIMA e Desenvolvimento: Visões para o Brasil 2030. *iCS - Clima e Sociedade*, 10 nov. 2021. Disponível em: <https://bit.ly/3uazklo>. Acesso em 19 abr. 2022.

COECKELBERGH, Mark. *AI Ethics*. Cambridge: MIT Press, 2020.

COECKELBERGH, Mark. Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. *Science and Engineering Ethics*, v. 26, n. 4, p. 2051-2068, out. 2019. Disponível em: <https://bit.ly/38439f5>. Acesso em: 5 abr. 2021.

COECKELBERGH, Mark. *The Political Philosophy of AI*. Cambridge: Polity Press, 2022.

COLEMAN, Flynn. *A Human Algorithm: How Artificial Intelligence is Redefining Who We Are*. Berkeley: Counterpoint, 2019.

CONSELHO NACIONAL DE JUSTIÇA. Resolução n.º 332, de 31 de agosto de 2020. Dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências. Brasília, 2020. Disponível em: <https://bit.ly/3KcTm4z>. Acesso em: 14 mar. 2022.

CONVERSACÕES – Dora Kaufman e Roberto Lent – Redes neurais cérebro humano. [S. l.: s. n.], 5 ago. 2020. 1 vídeo (1h 03min 32s). Publicado pelo canal TIDD PUC-SP. Disponível em: <https://bit.ly/3v4nKaO>. Acesso em: 11 abr. 2022.

COOK, Gary *et al.* Clicking Clean: Who is Winning the Race to Build a Green Internet? *Greenpeace*, jan. 2017. Disponível em: <https://bit.ly/3uaYVdU>. Acesso em: 14 mar. 2022.

CRAWFORD, Kate. *Atlas of AI*. New Haven; London: Yale University Press, 2021.

CRIAO PEREZ, Caroline. *Invisible Women: Exposing Data Bias in a World Designed for Men*. London: Vintage, 2021.

DARCY, Alison *et al.* Evidence of Human-Level Bonds Established With a Digital Conversational Agent: Cross-sectional, Retrospective Observational Study. *JMIR Formative Research*, v. 5, n. 5, 11 maio 2021. Disponível em: <https://bit.ly/3JYzVeJ>. Acesso em: 14 mar. 2022.

DAVIS, Kenrick. Welcome to China's latest "robot restaurant". *The World Economic Forum*, 01 jul. 2020. Disponível em: <https://bit.ly/361lQPF>. Acesso em: 11 abr. 2022.

DAVIS, Nicola. Scientists Create Online Games to Show Risks of ai Emotion Recognition. *The Guardian*, 4 abr. 2021. Disponível em: <https://bit.ly/3xALkPm>. Acesso em: 14 mar. 2022.

DE VYNCK, Gerrit. The U.S. says humans will always be in control of AI weapons: but the age of autonomous war is already here. *The Washington Post*, 7 jul. 2021. Disponível em: <https://wapo.st/3ju2mGt>. Acesso em: 11 abr. 2022.

DIGNUM, Virginia. *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Cham: Springer Netherlands, 2019.

DOCHERTY, Neil; FANNING, David. [S. l.: s. n.], 2019. 1 vídeo (1h 54min 16s). *In the Age of AI*. Publicado pelo canal Frontline PBS. Disponível em: https://youtu.be/5dZ_lvDgevk. Acesso em: 14 mar. 2022.

DOMINGOS, Pedro. *The Master Algorithm: How the Quest for the Ultimate Learning Machine will Remake our World*. Nova York: Basic Books, 2015.

DUHIGG, Charles. How Companies Learn Your Secrets. *The New York Times*, 16 fev. 2012. Disponível em: <https://nyti.ms/3Kg5GRn>. Acesso em: 14 mar. 2022; *O poder do hábito*. Trad. Rafael Mantovani. Rio de Janeiro: Objetiva, 2012.

ETHICS and Governance of ai at Berkman Klein: Report on Impact, 2017-2019. Berkman Klein Center, 11 out. 2019. Disponível em: <https://bit.ly/3DJrsdJ>. Acesso em: 14 mar. 2022.

EUBANKS, Virginia. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. Nova York: St. Martin's Press, 2018.

EUROPEAN COMMISSION, 18 abr. 2019. Disponível em: <https://bit.ly/3r4DXLZ>. Acesso em: 14 de mar. 2022.

EZRACHI, Ariel; STUCKE, Maurice E. *Virtual Competition: The promise and perils of the algorithm-driven economy*. Cambridge: Harvard University Press, 2016.

FERNANDES, Daniela. "Gigantes da web são novo Estado", diz Pierre Lévy. *Valor Econômico*, 23 out. 2020. Disponível em: <http://glo.bo/35ITFVz>. Acesso em: 14 mar. 2022.

FFHC Debate: capitalismo de vigilância e democracia, com Shoshana Zuboff. *Fundação FHC*, 9 dez. 2021. Disponível em: <https://bit.ly/3r7oHy2> Acesso em: 14 mar. 2022.

FILGUEIRAS, Fernando; ALMEIDA, Virgílio. *Governance for the Digital World: Neither More State nor More Market*. London: Palgrave Macmillan, 2021.

FINN, Ed. *What Algorithms Want: Imagination in the Age of Computing*. Cambridge: MIT Press, 2017.

FIRST Draft of Recommendation on Ethics of Artificial Intelligence. *Unesco*, 7 set. 2020. Disponível em: <https://bit.ly/3x8l4LY>. Acesso em: 14 mar. 2022.

FLORIDI, Luciano *et al.* An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, v. 28, n. 4, p. 689-707, nov. 2018. Disponível em: <https://bit.ly/3uPhsvC>. Acesso em: 4 abr. 2021.

FLORIDI, Luciano *et al.* The ethics of algorithms: Mapping the debate. *Big Data & Society*, v. 3, n. 2, p. 1-16, jul./dez. 2016. Disponível em: <https://bit.ly/3uVL0c5>. Acesso em: 4 abr. 2021.

FLORIDI, Luciano. AI and Its New Winter: from Myths to Realities. *Philosophy and Technology*, v. 33, 2020. Disponível em: <https://bit.ly/3j65RD1>. Acesso em: 14 mar. 2022.

FLORIDI, Luciano. *et al.* How to Design AI for Social Good: Seven Essential Factors. *Science and*

Engineering Ethics, v. 26, n. 3, p. 1771–1796, abr. 2020. Disponível em: <https://bit.ly/3vAM2JY>. Acesso em 06 abr. 2022.

FLORIDI, Luciano. The End of an Era: from Self-Regulation to Hard Law for the Digital Industry. *Philosophy and Technology*, v. 34, 2021. Disponível em: <https://bit.ly/3r4gg6G>. Acesso em: 14 mar. 2022.

FLORIDI, Luciano; COWLS, Josh. A United Framework of Five Principles for AI in Society. *Harvard Data Science Review*, v. 1, n. 1, p. 1-17, set. 2019. Disponível em: <https://bit.ly/3MjBo0z>. Acesso em: 4 abr. 2021.

FLORIDI, Luciano; SANDERS, John. W. On the Morality of Artificial Agents. *Minds and Machines*, v. 14, n. 3, p. 349–379, ago. 2004.

FOGG, Brian. Persuasive computers: perspectives and research directions. In: KARAT, Clare-Marie. *CHI '98: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY: Addison-Wesley Publishing Co., 1998. p. 225-232. Disponível em: <https://bit.ly/3DGjeTX>. Acesso em: 2 abr. 2022.

FORD, Martin. *Architects of Intelligence: the truth about AI from the people building it*. Birmingham: Packt, 2018.

FORSTER, E. M. *A máquina parou*. Org. e trad. Teixeira Coelho. São Paulo: Itaú Cultural; Iluminuras, 2018.

FORTIM, Ivelise; SPRITZER, Daniel Tornaim; LIMA, Maria Thereza Alencar. *Games viciam: fato ou ficção?* São Paulo: Estação das Letras e Cores, 2020.

FRANKLIN, Peter. How AI will terminate us. *Unherd*, 11 jun. 2019. Disponível em: <https://bit.ly/3LHTVUi>. Acesso em: 2 abr. 2022.

FREEMAN, Joshua. *Mastodontes: a história da fábrica e a construção do mundo moderno*. São Paulo: Todavia, 2019.

FREY, Carl Benedikt; OSBORNE, Michael A. The Future of Employment: How Susceptible are Jobs to Computerisation? *Oxford Martin School*, 01 set. 2013. Disponível em: <https://bit.ly/3uvSja8>. Acesso em: 11 abr. 2022.

FRIEDMAN, Batya; NISSENBAUM, Helen. Bias in Computer Systems. *ACM Transactions on Information Systems (TOIS)*, v. 14, n. 3, p. 330-347, jul. 1996. Disponível em: <https://bit.ly/3OjKkoG>. Acesso em 07 abr. 2022.

FRISCHMANN, Brett; SELINGER, Evan. *Re-Engineering Humanity*. Cambridge: Cambridge University Press, 2018.

FRY, Hannah. *Hello World: Being Human in the Age of Algorithms*. Nova York: W. W. Norton & Company, 2019.

GATHERING Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report. *Stanford University*, set. 2021. Disponível em: <https://stanford.io/3LXzvXu>. Acesso em: 11 abr. 2022.

GEBRU, Timnit. Race and Gender. In: DUBBER, Markus D.; PASQUALE, Frank; DAS, Sunit. (Eds.). *Oxford Handbook of Ethics of AI*. Nova York: Oxford University Press, 2020.

GERRISH, Sean. *How Smart Machine Think*. Cambridge: MIT Press, 2018.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. *Deep Learning*. Cambridge: MIT Press, 2016.

GROTHER, Patrick; NGAN, Mei; HANAOKA, Kayee. Face Recognition Vendor Test (frvt). *National Institute of Standards and Technology, U. S. Department of Commerce*, [s.d.]. Disponível em: doi.org/10.6028/NIST.IR.8280. Acesso em: 14 mar. 2022.

GUNKEL, David. *The Machine Question: Critical perspectives on AI, Robots, and Ethics*. Cambridge: MIT Press, 2012.

GUSTIN, Sam. IBM'S Watson Supercomputer Wins Practice Jeopardy Round. *Wired*, 13 jan. 2011. Disponível em: <https://bit.ly/37jcWNE>. Acesso em: 14 mar. 2022.

HALO Privacy Whitepaper. Disponível em: <https://bit.ly/36QByxK>. Acesso em: 14 mar. 2022.

HAN, Byung-Chul. (2017). *Sociedade da transparência*. Trad. Enio Paulo Giachini. Petrópolis: Editora Vozes, 2019.

HAN, Byung-Chul. O coronavírus de hoje e o mundo de amanhã. *El País*, 22 mar. 2020. Disponível em: <https://bit.ly/3ud57lZ>. Acesso em: 14 mar. 2022.

HAO, Karen. Intelligent Machines. This is how AI bias really happens: and why it's so hard to fix. *MIT Technology Review*, 4 fev. 2019. Disponível em: <https://bit.ly/3jT16wQ>. Acesso em: 7 abr. 2021.

HARARI, Yuval Noah. *Homo Deus, uma breve história do amanhã*. São Paulo: Companhia das Letras, 2015.

HARTLEY, Scott. *O fuzzy e o techie: por que as ciências humanas vão dominar o mundo digital*. São Paulo: Bei, 2017.

HARWELL, Drew. The Accent Gap. *The Washington Post*, 19 jul. 2018. Disponível em: <https://wapo.st/3Jcg2R3>. Acesso em: 14 mar. 2022.

HASHIGUCHI, Tiago Cravo Oliveira. Artificial intelligence in health still needs human intelligence: The COVID-19 crisis has shown that AI can deliver benefits, but it has also exposed its limits, often related to having the right data. *Health Policy Analyst*, OECD, 15 set. 2021. Disponível em: <https://bit.ly/3jU38gi>. Acesso em: 19 abr. 2022.

HASSE, Alexa *et al.* Youth and Artificial Intelligence: Where We Stand. *Berkman Klein Center for Internet and Society*, 2019. Disponível em: <https://bit.ly/37j39ao>. Acesso em: 14 mar. 2022.

HILL, Kashmir. The Secretive Company That Might End Privacy as We Know It. *The New York Times*, 18 jan. 2020. Disponível em: <https://nyti.ms/3j5Ftck>. Acesso em: 14 mar. 2022.

HLEGAI [High Level Expert Group on Artificial Intelligence]. Ethics Guidelines for Trustworthy AI. *European Commission*, 18 abr. 2019. Disponível em: <https://bit.ly/3r4DXLZ>. Acesso em: 14 de mar. 2022.

HO, Daniel *et al.* How regulators can get facial recognition technology right. *Brookings TechStream*, 7 nov. 2020. Disponível em: <https://brook.gs/37pnoE3>. Acesso em: 10 ago. 2021.

HODENT, Celia. Understanding the Success of Fortnite: A UX & Psychology Perspective. *Brains, UX & Games*, 22 jul. 2019. Disponível em: <https://bit.ly/3DXtzuV>. Acesso em: 14 mar. 2022; *The Gamer's Brain: How Neuroscience and UX Can Impact Video Game Design*. Boca Raton: Taylor & Francis, 2018.

HUSSAIN, Waheed. The Common Good. *Stanford Encyclopedia of Philosophy*. Disponível em: <https://stanford.io/3KdRaK1>. Acesso em: 14 mar. 2022.

HUTSON, Matthew. It's Too Easy to Hide Bias in Deep-Learning Systems. *IEEE Spectrum*, 19 jan. 2021. Disponível em: <https://bit.ly/3r63kgj>. Acesso em: 14 mar. 2022.

IANSTITI, Marco; LAKHANI, Karim R. Competing in the Age of AI. *Harvard Business Review*, jan./fev. 2020. Disponível em: <https://bit.ly/3jSAa0h>. Acesso em: 02 jan. 2020.

ISHIGURO, Kazuo. *Klara e o Sol*. Trad. Ana Guadalupe. São Paulo: Companhia das Letras, 2021.

JOHNSON, Bobbie; LICHFIELD, Gideon. Hey Google, Sorry You Lost Your Ethics Council, so We Made One for You. *MIT Technology Review*, 6 abr. 2019. Disponível em: <https://bit.ly/3NNbbcz>. Acesso em: 14 mar. 2022.

JOHNSON, Khari. DarwinAI Wants to Help Identify Coronavirus in X-rays, but Radiologists Aren't Convinced. *VentureBeat*, 25 mar. 2020. Disponível em: <https://bit.ly/3r5wLiK>. Acesso em: 14 mar. 2022.

JOHNSON, Khari. How People are Using AI to Detect and Fight the Coronavirus. *VentureBeat*, 3 mar. 2020. Disponível em: <https://bit.ly/3Jdm97q>. Acesso em: 14 mar. 2022.

JOHNSON, Khari. AI Weekly: A Deep Learning Pioneer's Teachable Moment on AI Bias. *VentureBeat*, 26 jun. 2020. Disponível em: <https://bit.ly/3jdfDmA>. Acesso em: 14 mar. 2022.

JUDEA, Pearl. *Causality: Models, Reasoning, and Inference*. Nova York: Cambridge University Press, 2000.

KAHNEMAN, Daniel; SIBONY, Olivier; SUNSTEIN, Cass R. *Noise: A Flaw in Human Judgement*. Nova York: Little, Brown Spark, 2021.

KALAJIAN, Rob. How Many People Play Fortnite in 2021?. *Sportskeeda*, 27 jan. 2021. Disponível em: <https://bit.ly/3OnXgd1>. Acesso em: 14 mar. 2022.

KALLENBORN, Zachary. Russia may have used a killer robot in Ukraine. Now what?. *Bulletin of the Atomic Scientists*, 15 mar. 2022. Disponível em: <https://bit.ly/3v4XME6>. Acesso em: 11 abr. 2022.

KALLENBORN, Zachary. Russia may have used a killer robot in Ukraine. Now what?. *Bulletin of the Atomic Scientists*, 15 mar. 2022. Disponível em: <https://bit.ly/3v4XME6>. Acesso em: 11 abr. 2022.

KAUFMAN, Dora. *A inteligência artificial irá suplantará a inteligência humana?*. São Paulo: Estação das Letras e Cores, 2019.

KAUFMAN, Dora. Inteligência Artificial e os desafios éticos: a restrita aplicabilidade dos princípios gerais para nortear o ecossistema de IA. *PAULUS: Revista de Comunicação da FAPCOM*, São Paulo, v. 5, n. 9, p. 73-84, jan./jul. 2021a. Disponível em: <https://doi.org/10.31657/rcp.v5i9.453>. Acesso em 06 abr. 2022.

KAUFMAN, Dora. Resenha do livro *Ethics of Artificial Intelligence*, de Matthew Liao. *Revista Digital de Tecnologias Cognitivas – TECCOGS*, n. 21, p. 157-163, jan./jun. 2020. Disponível em: <https://bit.ly/3NQ4GWj>. Acesso em: 14 mar. 2022.

KAUFMAN, Dora. Um projeto de futuro. *Piauí*, 22 out. 2021. Disponível em: <https://bit.ly/36SUw6S>. Acesso em: 14 mar. 2022.

KELLY, Kevin. *The Inevitable: understanding the 12 technological forces that will shape our future*. Nova York: Viking, 2016.

KELLY, Kevin. *What Technology Wants*. Nova York: Viking Press, 2010.

KENT, Chloe. Mental health chatbots might do better when they don't try to act human. *Medical Device Network*, 21 maio 2021. Disponível em: <https://bit.ly/3LI2DSj>. Acesso em: 2 abr. 2022.

KHAN, Lina M. Amazon's Antitrust Paradox. *Yale Law Journal*, v. 126, 2016. Disponível em: <https://bit.ly/372cVhk>. Acesso em: 14 mar. 2022.

KHARPAL, Arjun. Coronavirus could be a “catalyst” for China to boots its mass surveillance machine, experts say. *CNBC*, 24 fev. 2020.

KING, Brett. *Bank 4.0: Banking Everywhere, Never at a Bank*. Chichester: John Wiley & Sons, 2019.

KIRKPATRICK, David. The Facebook Defect. *Time*, 12 abr. 2018. Disponível em: <https://time.com/5237458/the-facebook-defect/>. Acesso em: 14 mar. 2022.

KNIGHT, Will. Russia's Killer Drone in Ukraine Raises Fears About AI in Warfare. *Wired*, 17 mar. 2022. Disponível em: <https://bit.ly/3O3huJ6>. Acesso em: 11 abr. 2022.

KOBIE, Nicole. Reuters Is Taking a Big Gamble on AI-supported Journalism. *Wired*, 10 mar. 2018. Disponível em: <https://bit.ly/3DHbv81>. Acesso em: 14 mar. 2022.

KOENECKE, Allison *et al.* Racial Disparities in Automated Speech Recognition. *pnas*, 7 abr. 2020. Disponível em: pnas.org/content/117/14/7684. Acesso em: 14 mar. 2022.

KRAFT-BUCHMAN, Caitlin. Chapter 1: We Shape Our Tools, and Thereafter Our Tools Shape Us. *Feminist AI*, 22 jun. 2021. Disponível em: <https://bit.ly/3rBU9op>. Acesso em 06 abr. 2022.

KRIZHEVSHKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, v. 25, p. 1-9, 2012. Disponível em: <https://bit.ly/3jSuRd6>. Acesso em: 12 jan. 2019.

KROGSTAD, Jens Manuel; LOPEZ, Mark Hugo. Black Voter Turnout Fell in 2016, even as a Record Number of Americans Cast Ballots. *Pew Research Center*, 12 maio 2017. Disponível em: <https://pewrsr.ch/37mcq1G>. Acesso em: 14 mar. 2022.

KURZWEIL, Ray. (2005). *A singularidade está próxima: quando os humanos transcendem a biologia*. Trad. Ana Goldberger. São Paulo: Itaú Cultural; Iluminuras, 2018.

LANDHUIS, Esther. Neuroscience: Big Brain, Big Data. *Nature*, v. 541, p. 559-561, 26 jan. 2017. Disponível em: doi.org/10.1038/541559a. Acesso em: 14 mar. 2022.

LEARNED-MILLER, Erik *et al.* Facial Recognition Technologies in the Wild: A Call for a Federal Office. *Algorithmic Justice League*, 2020. Disponível em: <https://bit.ly/3jSAEn7>. Acesso em: 15 set. 2021.

LEE, Kai-Fu. A Human Blueprint for AI Coexistence. In: BRAUN, Joachim von; REICHBERG, Gregory

M.; ARCHER, Margaret S.; SORONDO, Marcelo Sánchez. *Robotics, AI, and Humanity: Science, Ethics, and Policy*. Nova York: Springer, 2021. p. 263. Disponível em: <https://bit.ly/38vfMzN>. Acesso em: 11 abr. 2022.

LEE, Kai-Fu. *AI Superpowers: China, Silicon Valley, and the New World Order*. New York: Harper Business, 2018. Ed. bras: *Inteligência artificial*. Trad. Marcelo Barbão. Rio de Janeiro: Globo Livros, 2019.

LEE, Kai-Fu. *Inteligência artificial: como os robôs estão mudando a forma como amamos, nos relacionamos, trabalhamos e vivemos*. Tradução de Marcelo Barbão. Rio de Janeiro: Globo Livros, 2019.

LEE, Kai-Fu; QIU-FAN, Chen. *AI 2041: Ten Visions for our Future*. Nova York: Currency; Penguin Random House, 2021.

LEMBKE, Anna. *Nação dopamina: por que o excesso de prazer está nos deixando infelizes e o que podemos fazer para mudar*. São Paulo: Vestígio, 2022.

LEMONS, Ronaldo. Estratégia de ia brasileira é patética. *Folha de S.Paulo*, 11 abr. 2021. Disponível em: <https://bit.ly/3NQc43Z>. Acesso em: 14 mar. 2022.

LENT, Roberto. *O cérebro aprendiz: neuroplasticidade e educação*. São Paulo: Atheneu, 2018.

LESLIE, David. Understanding bias in facial recognition technologies: an explainer. *The Alan Turing Institute*, 2020. Disponível em: <https://doi.org/10.5281/zenodo.4050457>. Acesso em: 17 set. 2021.

LIAO, S. Matthew. *Ethics of Artificial Intelligence*. Nova York: Oxford University Press, 2020.

LIMA, Lanna *et al.* Empirical Analysis of Bias in Voice-based Personal Assistants. In: LIU, Lin; WHITE, Ryan. *www'19: Companion Proceedings of The 2019 World Wide Web Conference*, maio 2019. Disponível em: <https://bit.ly/3j6zjZl>. Acesso em: 14 mar. 2022.

LIU, Xiao; MERRITT, Jeff. Shaping the Future of the Internet of Bodies: New Challenges of Technology Governance. Briefing Paper. *The World Economic Forum*, jul. 2020. Disponível em: <https://bit.ly/3ukWUMF>. Acesso em: 14 mar. 2022.

LOVELOCK, James. *Novacene: The Coming Age of Hyperintelligence*. London: Penguin, 2019.

LUNDBERG, Scott. M.; LEE, Su-In. A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, v. 30, p. 4765-4774, 2017. Disponível em: <https://bit.ly/3uXPHT0>. Acesso em: 15 set. 2021.

MAIA FILHO, Mamede Said; JUNQUILHO, Tainá Aguiar. Projeto Victor: perspectivas de aplicação da inteligência artificial ao direito. *Revista de Direitos e Garantias Fundamentais*, v. 19, n. 3, 2018. Disponível em: <https://bit.ly/3LGgqcb>. Acesso em: 14 mar. 2022.

MARCUS, Gary; DAVIS, Ernest. *Rebooting AI: Building Artificial Intelligence We Can Trust*. Nova York: Vintage Books; Penguin Random House, 2020.

MARR, Bernard. Artificial Intelligence Can Now Write Amazing Content: What Does That Mean for Humans? *Forbes*, 29 mar. 2019. Disponível em: <https://bit.ly/3uTXlwd>. Acesso em: 14 mar. 2022.

MATWYSHYN, Andrea M. The “Internet of Bodies” Is Here. Are Courts and Regulators Ready? *The Wall Street Journal*, 12 nov. 2018. Disponível em: <https://on.wsj.com/36Yt739>. Acesso em: 14 mar. 2022.

MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Nova York: Houghton Mifflin Harcourt, 2013.

MAYER-SCHÖNBERGER, Viktor; RAMGE, Thomas. *Reinventing Capitalism in the Age of Big Data*. Nova York: Basic Books, 2018.

McCARTHY, John *et al.* A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. *Stanford*, 1955. Disponível em: <https://stanford.io/3MhSkEY>. Acesso em 1 nov. 2021.

McDERMOTT, Drew. Artificial Intelligence Meets Natural Stupidity. *ACM SIGART Bulletin*, v. 57, abr. 1976. Disponível em: <https://bit.ly/38zQoct>. Acesso em: 14 mar. 2022.

MCEWAN, Ian. *Máquinas como eu*. Trad. Jorio Dauster. São Paulo: Companhia das Letras, 2019.

MCQUAID, John. The True Cost of the Massive Datasets A.I. Uses to Learn. *Slate*, 3 nov. 2021. Disponível em: <https://bit.ly/3NP7KBY>. Acesso em: 14 mar. 2022.

MELLO, Michelle M.; WANG, C. Jason. Ethics and Governance for Digital Disease Surveillance. *Science*, 29 maio 2020. Disponível em: <https://bit.ly/3DFVDmh>. Acesso em: 14 mar. 2022.

MINSKY, Marvin. Communication with Alien Intelligence. In: REGIS, Edward (Ed.). *Extraterrestrials: Science and Alien Intelligence*. Cambridge; Nova York: Cambridge University Press, 1985. Disponível em: <https://bit.ly/3vsTMNN>. Acesso em: 15 set. 2021.

MINSKY, Marvin. Steps Toward Artificial Intelligence. *Proceedings of the Institute of Radio Engineers*, v. 49, n. 1, p. 8-30, 1961.

MINSKY, Marvin; PAPPET, Seymour. *Perceptrons: An Introduction to Computational Geometry*. Cambridge: MIT Press, 1969.

MITCHELL, Melanie. *Artificial Intelligence: A Guide for Thinking Humans*. Nova York: Farrar, Straus and Giroux, 2019.

MITCHELL, Melanie. Why AI Is Harder Than We Think. *ArXiv*, 28 abr. 2021. Disponível em: arxiv.org/abs/2104.12871. Acesso em: 14 mar. 2022.

MÖKANDER, Jakob; FLORIDI, Luciano. Ethics-Based Auditing to Develop Trustworthy AI. In: *Minds and Machines*, v. 31, n. 2, p. 323-327, jun. 2021. Disponível em: <https://bit.ly/36YnT7x>. Acesso em: 4 abr. 2021.

MORIN, Edgar. Um festival de incerteza. Trad. Edgar Carvalho e Fagner França. *Instituto Humanitas Unisinos*, 9 jun. 2020. Disponível em: <https://bit.ly/3jdeXO4>. Acesso em: 14 mar. 2022.

MOROZOV, Evgeny. *Big Tech: a ascensão dos dados e a morte da política*. São Paulo: Ubu, 2018.

MOZUR, Paul; ZHONG, Raymond; KROLIK, Aaron. In Coronavirus Fight, China Gives Citizens a Color Code, With Red Flags. *The New York Times*, 01 mar. 2020, disponível em: <https://nyti.ms/3DMGvDM>. Acesso em: 14 mar. 2022.

MULAS-GRANADOS, Carlos *et al.* Automation, Skills and the Future of Work: What do Workers Think? *Imf Working Papers*, dez. 2019. Disponível em: <https://bit.ly/3uVIMsa>. Acesso em: 14 mar. 2022.

MUSSA, Adriano. *Inteligência artificial, mitos e verdades*. São Paulo: Saint Paul, 2020.

NEUMAN, W. Russell. *The Digital Difference: Media Technology and the Theory of Communication Effects*. Cambridge: Harvard University Press, 2016.

NEWMAN, Nic. Journalism, Media, and Technology Trends and Predictions 2021. *Reuters Institute*, 7 jan. 2021. Disponível em: <https://bit.ly/3NPX7PC>. Acesso em: 14 mar. 2022.

NISSON, Nils J. *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. Nova York: Cambridge University Press, 2010.

O'MEARA, Sarah. AI Researchers in China Want to Keep the Global-sharing Culture Alive. *Nature*, v. 569, 29 maio 2019. Disponível em: <https://go.nature.com/3M0OrnR>. Acesso em: 14 mar. 2022.

O'NEIL, Cathy. (2016). *Algoritmos de destruição em massa: como o Big Data aumenta a desigualdade e ameaça a democracia*. Trad. Rafael Abraham. Santo André: Rua do Sabão, 2020.

O'NEIL, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown, 2016.

OECD. Recommendation of the Council on Artificial Intelligence. *OECD Legal Instruments*, 21 maio 2019. Disponível em: <https://bit.ly/3r6a3H5>. Acesso em 19 abr. 2022.

OXFAM. Inequality Kills. *Oxfam*, 16, jan. 2022. Disponível em: <https://bit.ly/3JaRSXa>. Acesso em: 14 mar. 2022.

PARISER, Eli. (2011). *O filtro invisível: o que a internet está escondendo de você* Trad. Diego Alfaro. Rio de Janeiro: Zahar, 2012.

PASQUALE, Frank. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge: Harvard University Press, 2015.

PEARL, Judea; MACKENZIE, Dana. *The Book of Why: The New Science of Cause and Effect*. Lavergne: Ingram Publisher, 2017.

PENTLAND, Alex. *Social Physics: How Social Networks Can Make Us Smarter*. Nova York: Penguin Books, 2015.

PEOPLE + AI Guidebook. Disponível em: <https://pair.withgoogle.com/guidebook>. Acesso em: 01 abr. 2022.

PEREZ-CRIADO, Caroline. *Invisible Women: Data Bias in a World Designed for Men*. Nova York: Abrams Press, 2021.

POWERING the Cloud: How China's Internet Industry Can Shift to Renewable Energy. *Greenpeace*, 2019. Disponível em: <https://bit.ly/3uYrBGd>. Acesso em: 14 mar. 2022.

PROPOSAL for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. *Comissão Europeia*, Bruxelas, 21 abr. 2021. Disponível em: <https://bit.ly/3DGFT2q>. Acesso em: 14 mar. 2022.

RESEARCHERS Will Deploy AI to Better Understand Coronavirus. *Wired*, 17 mar. 2020. Disponível em: <https://bit.ly/3713DSN>. Acesso em: 14 mar. 2022.

RITTNER, Daniel. “A economia global mudou e ficará mais fragmentada”, diz professor de Harvard. *Valor econômico*, 15 mar. 2022. Disponível em: <http://glo.bo/377PeV2>. Acesso em: 11 abr. 2022.

ROBERTS, Huw; COWLS, Josh; MORLEY, Jessica *et al.* The Chinese Approach to Artificial Intelligence: An Analysis of Policy, Ethics, and Regulation. *AI & Society*, v. 36, 2021. Disponível em: <https://bit.ly/35IW16R>. Acesso em: 14 mar. 2022.

ROLNICK, David *et al.* Tackling Climate Change with Machine Learning. *ArXiv*, 5 nov. 2019. Disponível em: arxiv.org/pdf/1906.05433.pdf. Acesso em: 14 mar. 2022.

ROSS, Alec. *The Industries of the Future*. New York: Simon & Schuster, 2016.

RUSSELL, Stuart. *Human Compatible: Artificial Intelligence and the Problem of Control*. New York: Viking; Penguin, 2019.

RUSSELL, Stuart. *Inteligência artificial a nosso favor: como manter o controle sobre a tecnologia*. São Paulo: Companhia das Letras, 2021.

RUSSELL, Stuart J.; NORVING, Peter. *Artificial Intelligence: A Modern Approach*. 3. ed. New Jersey: Prentice Hall, 2009.

SALOMÃO, Luís Felipe (Org.). *Inteligência artificial: tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário Brasileiro*. fgv Conhecimento; Centro de Inovação, Administração e Pesquisa do Judiciário, 2020. Disponível em: <https://bit.ly/3DHmpdX>. Acesso em: 14 mar. 2022.

SANTAELLA, Lucia. *Humanos Hiper-Híbridos: Linguagens e cultura na segunda era da internet*. São Paulo: Paulus, 2022.

SCHNEIDER, Susan. *Artificial You: AI and the Future of Your Mind*. Princeton: Princeton University Press, 2019.

SEJNOWSKI, Terrence J. *The Deep Learning Revolution: Artificial Intelligence Meets Human Intelligence*. Cambridge: MIT Press, 2018.

SEN, Amartya. *Sobre ética e economia*. Trad. Laura Teixeira Motta. São Paulo: Companhia das Letras, 1999.

SHEAD, Sam. Google DeepMind is Edging Towards a 3-0 Victory against World Go Champion Ke Jie. *Insider*, 25 maio 2017. Disponível em: <https://bit.ly/3DIyVK8>. Acesso em: 14 mar. 2022.

SHENG, Emily *et al.* The Woman Worked as a Babysitter: On Biases in Language Generation. In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING (EMNLP), 2019, p. 1-8. Disponível em: <https://arxiv.org/abs/1909.01326>. Acesso em: 19 mar. 2022.

SHERMAN, Jason. Russia-Ukraine Conflict Prompted U.S. to Develop Autonomous Drone Swarms, 1.000-Mile Cannon. *Scientific American*, 14 fev. 2022. Disponível em: <https://bit.ly/3uxfuRq>. Acesso em: 11 abr. 2022.

SHERMAN, Jason. Russia-Ukraine Conflict Prompted U.S. to Develop Autonomous Drone Swarms,

1.000-Mile Cannon. *Scientific American*, 14 fev. 2022. Disponível em: <https://bit.ly/3uxfuRq>. Acesso em: 11 abr. 2022.

SHIRKY, Clay. (2010). *A cultura da participação: criatividade e generosidade no mundo conectado*. Trad. Celina Portocarrero. Rio de Janeiro: Zahar, 2011.

SHLADOVER, Steven. “Self-Driving” Cars Begin to Emerge from a Cloud of Hype. *Scientific American*, 25 set. 2021. Disponível em: <https://bit.ly/3uc2YXF>. Acesso em: 14 mar. 2022.

SILBERG, Jake; MANYIKA, James. Como lidar com vieses na inteligência artificial (e nos seres humanos). *McKinsey Global Institute*, 6 jun. 2019. Disponível em: <https://bit.ly/36SZXmi>. Acesso em: 14 mar. 2022.

SIMONDON, Gilbert. (1958). *Do modo de existência dos objetos técnicos*. trad. Vera Ribeiro. Rio de Janeiro: Contraponto, 2020.

SIMONS, Josh; GHOSH, Dipayan. Utilities for Democracy: Why and How the Algorithmic Infrastructure of Facebook and Google Must Be Regulated. *Brookings*, ago. 2020. Disponível em: <https://brook.gs/3JbXkZG>. Acesso em: 14 mar. 2022.

SKINNER, Burrhus Frederic. *Science and Human Behavior*. Nova York: Macmillan, 1953.

SKJUVE, Marita *et al.* My Chatbot Companion: A Study of Human-Chatbot Relationship. *International Journal of Human-Computer Studies*, v. 149, maio 2021. Disponível em: <https://bit.ly/3NQgsjG>. Acesso em: 14 mar. 2022.

SMITH, Brad; SHUM, Harry. Foreword: The Future Computed. In: *The Future Computed: Artificial Intelligence and its Role in Society*. Washington: Microsoft, 2018.

SMITH, Gary. *The AI Delusion*. Oxford: Oxford University Press, 2018.

SOUSA, Joana Silva *et al.* *Employment in Crisis: The Path to Better Jobs in a Post-COVID-19 Latin America*. Washington, D.C.: World Bank Group, 2021. Disponível em: <https://bit.ly/3xjM4s7>. Acesso em: 11 abr. 2022.

SPIEGELHALTER, David. *The Art of Statistics Learning from Data*. Londres: Pelican Books, 2019.

STEPHENSON, Neal. *Snow Crash*. Trad. Fábio Fernandes. São Paulo: Aleph, 2015.

STRUBELL, Emma; GANESH, Ananya; MCCALLUM, Andrew. Energy and Policy Considerations for Deep Learning in NPL. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, jan. 2019. Disponível em: <https://bit.ly/3v13wil>. Acesso em: 14 mar. 2022.

SUNSTEIN, Cass R. *#republic: Divided Democracy in the Age of Social Media*. Princeton: Princeton University Press, 2017.

THE AI Index Report: Measuring Trends in Artificial Intelligence. *Stanford Institute for Human-Centered Artificial Intelligence*, 2021. Disponível em: <https://stanford.io/3Kg8C0l>. Acesso em: 14 mar. 2022.

TOEWS, Rob. Deep Learning's Carbon Emissions Problem. *Forbes*, 17 jun. 2020. Disponível em: <https://bit.ly/3j8EOag>. Acesso em: 14 mar. 2022.

TOPOL, Eric. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. Nova York: Basic Books, 2019.

TRICKEY, Erick. The Law and the Digital World. *Harvard Law Today*, 3 abr. 2019. Disponível em: <https://bit.ly/3u814Hs>. Acesso em: 14 mar. 2022.

TURING, Alan M. Computing Machinery and Intelligence. *Mind*, v. 59, n. 236, p. 433-460, out. 1950. Disponível em: <https://doi.org/10.1093/mind/LIX.236.433>. Acesso em: 07 abr. 2022.

TURING, Alan. Can Digital Computers Think? In: COPELAND, B. Jack (Ed.). *The Essential Turing*. New York: Oxford University Press, 2004.

TURKLE, Sherry. *Alone Together: Why We Expect More from Technology and Less from Each Other*. New York: Basic Books, 2017

U.S. OPIOID Dispensing Rate Maps. Disponível em: cdc.gov/drugoverdose/rxrate-maps/index.html. Acesso em: 14 mar. 2022.

UNDERWOOD, Corinna. Automated Journalism: AI applications at *New York Times*, Reuters and Other Media Giants. *Emerj*, 17 nov. 2019. Disponível em: <https://bit.ly/3jaV0HM>. Acesso em: 14 mar. 2022.

UNODC – United Nations Office on Drugs and Crime. Disponível em: <https://bit.ly/3KYDRxr>. Acesso em: 14 mar. 2022.

VARGA, George. How Peter Jackson shifted the dour context of “Let It Be” into a fuller, more human “Get Back”. *The Washington Post*, 25 nov. 2021. Disponível em: <https://bit.ly/3uv2lZq>. Acesso em: 11 abr. 2022.

VELASQUEZ-MANOFF, Moises. The Brains Implant That Could Change Humanity. *The New York Times*, 28 ago. 2020. Disponível em: <https://nyti.ms/3NOUJZe>. Acesso em: 14 mar. 2022.

VILLANI, Cédric. For a Meaningful Artificial Intelligence: Towards a French and European Strategy. *AI for Humanity*, 2018. Disponível em: <https://bit.ly/383n3GS>. Acesso em: 10 abr. 2021.

VINUESA, R.; AZIZPOUR, H.; LEITE, I. *et al.* The Role of Artificial Intelligence in Achieving the Sustainable Development Goals. *Nature Communications*, 2020. Disponível em: <https://go.nature.com/3uVZ30e>. Acesso em: 14 mar. 2022.

WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. *International Data Privacy Law*, v. 7, n. 2, p. 76-99, jun. 2017. Disponível em: <https://doi.org/10.1093/idpl/ipy005>. Acesso em: 7 abr. 2022.

WALLACH, Wendell; ALLEN, Colin. *Moral Machines: Teaching Robots Right from Wrong*. Oxford: Oxford University Press, 2009

WEST, Mark; KRAUT, Rebecca; CHEW, Han Ei. I'd Blush, If I Could: Closing Gender Divides in Digital Skills Through Education. *Equals*, Unesco, 2019. Disponível em: <https://en.unesco.org/ld-blush-if-i-could>. Acesso em: 14 mar. 2022.

WEXLER, James *et al.* The What-If Tool: Interactive Probing of Machine Learning Models. *IEEE Transactions on Visualization and Computer Graphics*, v. 26, n. 1, p. 56-65, jan. 2020. Disponível em: <https://doi.org/10.1109/TVCG.2019.2934619>. Acesso em: 15 set. 2022.

WHITTLESTONE, Jess *et al.* Artificial Intelligence in a Crisis Needs Ethics With Urgency. *Nature Machine Intelligence*, 22 jun. 2020. Disponível em: nature.com/articles/s42256-020-0195-0. Acesso em: 14 mar. 2022.

WHO. Report of the WHO-China Joint Mission on Coronavirus Disease. *World Health Organization*, 16-24 fev. 2020. Disponível em: <https://bit.ly/3DQMuHG>. Acesso em: 14 mar. 2022.

WILDE, Oscar. *The Decay of Lying and Other Essays*. London: Penguin Books, 2020.

WOEBOT HEALTH. Disponível em: <https://woebothealth.com/about-us>. Acesso em: 2 abr. 2022.

WU, Tim. *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*. New York: Knopf, 2016.

WU, Tim. *The Curse of Bigness: Antitrust in the New Gilded Age*. New York: Columbia Global Reports, 2018.

WYNANTS, Laure *et al.* Prediction Models for Diagnosis and Prognosis of Covid-19 Infection: Systematic Review and Critical Appraisal. *The British Medical Journal*, 7 abr. 2020. Disponível em: <https://bit.ly/36Ylbiv>. Acesso em: 14 mar. 2022.

ZHANG, Baobao; DAFOE, Allan. Artificial Intelligence: American Attitudes and Trends. *Social Science Research Network*, 19 jan. 2019. Disponível em: <https://bit.ly/3xknjf5>. Acesso em: 11 abr. 2022.

ZIERAHN, Ulrich. Race Against the Machine? The Role of Technological Change for Employment, Wages and Inequality. *Israel Public Policy Institute*, 27 jan. 2022. Disponível em: <https://bit.ly/3KuCyG8>. Acesso em: 11 abr. 2022.

ZUBOFF, Shoshana. (2019). *A era do capitalismo de vigilância*. Trad. George Schlesinger. Rio de Janeiro: Intrínseca, 2021.

ZUBOFF, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Nova York: PublicAffairs, 2019.

PhD pela Carnegie Mellon University, professor titular da Politécnica e livre-docente da USP, diretor do Centro de Inteligência Artificial (C4AI) da USP/IBM/FAPESP.

The Guardian, 12 jan. 2010.

Essa história e muitos outros casos curiosos são relatados no livro *Machines Who Think*, de Pamela McCorduck (2004).

Vídeo da palestra disponível em: <https://bit.ly/3J9TmRl>. Acesso em: 5 abr. 2022.

RUSSELL, Stuart. *Human Compatible: Artificial Intelligence and the Problem of Control*. New York: Viking; Penguin, 2019.

BENJAMIN, Ruha. Assessing Risk, Automating Racism. *Science*, 25 out. 2019. Disponível em: <https://bit.ly/35GHhW2>. Acesso em: 14 mar. 2022.

No artigo “Um projeto de futuro”, ofereço exemplos ilustrativos da aplicação de ia em diversos setores de atividade. Ver revista *Piauí*, 22 out. 2021. Disponível em: <https://bit.ly/3LGt9LV>. Acesso em: 14 mar. 2022.

HARTLEY, Scott. *O fuzzy e o techie: por que as ciências humanas vão dominar o mundo digital*. São Paulo: Bei, 2017.

Disponível em: <https://bit.ly/3x3dZMz>. Acesso em: 14 mar. 2022.

Os 17 ensaios inéditos, produzidos por cientistas e filósofos, agrupados em quatro seções, abordam dilemas-chave para uma inteligência artificial ética, entre outros, os impactos da automação inteligente no mercado de trabalho; o viés contido nos dados que perpetuam os preconceitos da sociedade; a ética envolvida em aplicações como carros autônomos, sistemas de vigilância, armas autônomas; robôs sexuais; direitos e consciência da ia; e status moral. Numa perspectiva futura, os ensaios da terceira seção refletem sobre os riscos da “superinteligência”. Ver LIAO, S. Matthew. *Ethics of Artificial Intelligence*. Oxford: Oxford University Press, 2020; KAUFMAN, Dora. Resenha do livro *Ethics of Artificial Intelligence*, de Matthew Liao. *Revista Digital de Tecnologias Cognitivas – TECCOGS*, n. 21, p. 157-163, jan./jun. 2020. Disponível em: <https://bit.ly/3NQ4GWj>. Acesso em: 14 mar. 2022.

LENT, Roberto. *O cérebro aprendiz: neuroplasticidade e educação*. São Paulo: Atheneu, 2018.

LENT. *O cérebro aprendiz*.

BUTLER, Samuel. *Erewhon* [1872]. London: Penguin, 1970.

TURING, Alan. Can Digital Computers Think? In: COPELAND, B. Jack (Ed.). *The Essential Turing*. New York: Oxford University Press, 2004.

HARARI, Yuval Noah. *Homo Deus: uma breve história do amanhã* [2015]. Trad. Paulo Geiger. São Paulo: Companhia das Letras, 2016.

RUSSELL, Stuart. *Human Compatible: Artificial Intelligence and the Problem of Control*. New York: Viking; Penguin, 2019.

TURING, Alan. Computing Machinery and Intelligence. *Mind Magazine*, v. 59, n. 236, p. 433-60, 01 out. 1950.

MCCARTHY, J. *et al.* *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 1956. Disponível em: <https://bit.ly/3xIKNea>. Acesso em: 15 abr. 2022.

KURZWEIL, Ray. (2005). *A singularidade está próxima: quando os humanos transcendem a biologia*. Trad. Ana Goldberger. São Paulo: Itaú Cultural; Iluminuras, 2018.

FORD, Martin. *Architects of Intelligence: the truth about AI from the people building it*. Birmingham: Packt, 2018. p. 33.

Entrevista realizada no ciclo de debates Conversações, do TIDDigital, promovido pelo Programa de Estudos Pós-Graduados em Tecnologias da Inteligência e Design Digital da PUC-SP, em 5 de agosto de 2020. Disponível em: <https://youtu.be/0PrG5IHBLVk>. Acesso em: 14 mar. 2022.

FORD, Martin. *Architects of Intelligence: the truth about AI from the people building it*. Birmingham: Packt, 2018. p. 193.

PARISER, Eli. (2011). *O filtro invisível: o que a internet está escondendo de você*. Trad. Diego Alfaro. Rio de Janeiro: Zahar, 2012.

SUNSTEIN, Cass R. *#republic: Divided Democracy in the Age of Social Media*. Princeton: Princeton University Press, 2017.

ZUBOFF, Shoshana. (2019). *A era do capitalismo de vigilância*. Trad. George Schlesinger. Rio de Janeiro: Intrínseca, 2021.

People + AI Guidebook. Disponível em: <https://pair.withgoogle.com/guidebook>. Acesso em: 01 abr. 2022.

HUTSON, Matthew. It's Too Easy to Hide Bias in Deep-Learning Systems. *IEEE Spectrum*, 19 jan. 2021. Disponível em: <https://bit.ly/3r63kgj>. Acesso em: 14 mar. 2022.

Ver MCQUAID, John. The True Cost of the Massive Datasets A.I. Uses to Learn. *Slate*, 3 nov. 2021. Disponível em: <https://bit.ly/3NP7KBY>. Acesso em: 14 mar. 2022.

CHRISTIAN, Brian. *The Alignment Problem: Machine Learning and Human Values*. New York: W. W. Norton & Company, 2020.

MCQUAID. The True Cost of the Massive Datasets A.I. Uses to Learn.

CHOI, C. Q. 7 Revealing Ways AIs Fail: Neural Networks can be Disastrously Brittle, Forgetful, and Surprisingly Bad at Math. *IEEE Spectrum*, v. 58, n. 10, out. 2021. Disponível em: <https://bit.ly/3x3eLsX>. Acesso em: 14 mar. 2022.

Disponível em: <https://bit.ly/3JdWvPY>. Acesso em: 2 abr. 2022.

MULAS-GRANADOS, Carlos *et al.* Automation, Skills and the Future of Work: What do Workers Think? *Inf Working Papers*, dez. 2019. Disponível em: <https://bit.ly/3uVIMsa>. Acesso em: 14 mar. 2022.

BESS, Michael. *Our Grandchildren Redesigned: Life in the Bioengineered Society of the Near Future*. Boston: Beacon Press, 2015.

HARARI. *Homo Deus*.

LEE, Kai-Fu. *AI Superpowers: China, Silicon Valley, and the New World Order*. New York: Harper Business, 2018. Ed. bras: *Inteligência artificial*. Trad. Marcelo Barbão. Rio de Janeiro: Globo Livros, 2019.

FREEMAN, Joshua. *Mastodontes: a história da fábrica e a construção do mundo moderno*. São Paulo: Todavia, 2019.

HARTLEY, Scott. *O fuzzy e o techie: por que as ciências humanas vão dominar o mundo digital*. São Paulo: Bei, 2017.

Disponível em: <https://bit.ly/37gDBL9>. Acesso em: 14 mar. 2022.

NEWMAN, Nic. Journalism, Media, and Technology Trends and Predictions 2021. *Reuters Institute*, 7 jan. 2021. Disponível em: <https://bit.ly/3NPX7PC>. Acesso em: 14 mar. 2022.

Advertising Age.

UNDERWOOD, Corinna. Automated Journalism: AI applications at *New York Times*, Reuters and Other Media Giants. *Emerj*, 17 nov. 2019. Disponível em: <https://bit.ly/3jaV0HM>. Acesso em: 14 mar. 2022.

KOBIE, Nicole. Reuters is Taking a Big Gamble on ai-supported Journalism. *Wired*, 10 mar. 2018. Disponível em: <https://bit.ly/3DHbv81>. Acesso em: 14 mar. 2022.

MARR, Bernard. Artificial Intelligence Can Now Write Amazing Content: What Does That Mean for Humans? *Forbes*, 29 mar. 2019. Disponível em: <https://bit.ly/3uTXlwd>. Acesso em: 14 mar. 2022.

ZIERAHN, Ulrich. Race Against the Machine? The Role of Technological Change for Employment, Wages and Inequality. *Israel Public Policy Institute*, 27 jan. 2022. Disponível em: <https://bit.ly/3KuCyG8>. Acesso em: 11 abr. 2022.

AUTOR, David H.; LEVY, Frank; MURNANE, Richard J. The Skill Content of Recent Technological Change: An Empirical Exploration Get access Arrow. *The Quarterly Journal of Economics*, v. 118, issue 4, p. 1279-1333, 01 nov. 2003. Disponível em: <https://bit.ly/37365Zf>. Acesso em: 11 abr. 2022.

BERG, Andrew; BUFFIE, Edward F.; ZANNA, Luis-Felipe. IMF Working Paper: Should We Fear the Robot Revolution? (The Correct Answer is Yes). *International Monetary Fund*, 2018. Disponível em: <https://bit.ly/3LRMOsk>. Acesso em: 11 abr. 2022.

LEE, Kai-Fu. A Human Blueprint for AI Coexistence. In: BRAUN, Joachim von; REICHBERG, Gregory M.; ARCHER, Margaret S.; SORONDO, Marcelo Sánchez (Eds.). *Robotics, AI, and Humanity: Science, Ethics, and Policy*. Nova York: Springer, 2021. p. 263. Disponível em: <https://bit.ly/38vfMzN>. Acesso em: 11 abr. 2022.

Disponível em: <https://bit.ly/3F8SNH9>. Acesso em 28 abr. 2022.

FREY, Carl Benedikt; OSBORNE, Michael A. The Future of Employment: How Susceptible are Jobs to Computerisation? *Oxford Martin School*, 01 set. 2013. Disponível em: <https://bit.ly/3uvSja8>. Acesso em: 11 abr. 2022.

RITTNER, Daniel. “A economia global mudou e ficará mais fragmentada”, diz professor de Harvard. *Valor econômico*, 15 mar. 2022. Disponível em: <http://glo.bo/377PeV2>. Acesso em: 11 abr. 2022.

SOUSA, Joana Silva et al. *Employment in Crisis: The Path to Better Jobs in a Post-COVID-19 Latin America*. Washington, D.C.: World Bank Group, 2021. Disponível em: <https://bit.ly/3xjM4s7>. Acesso em: 11 abr. 2022.

Para aprofundar esse debate, recomendo a leitura de *AI Ethics*, do filósofo Mark Coeckelbergh (MIT Press, 2020).

Disponível em: <https://bit.ly/3l22GNm>. Acesso em: 15 abr. 2022.

McEWAN, Ian. *Máquinas como eu*. Trad. Jorio Dauster. São Paulo: Companhia das Letras, 2019.

Disponível em: <https://bit.ly/3uYdKjb>. Acesso em: 2 abr. 2022.

HILL, Kashmir. The Secretive Company That Might End Privacy as We Know It. *The New York Times*, 18 jan. 2020. Disponível em: <https://nyti.ms/3j5Ftck>. Acesso em: 14 mar. 2022.

WHITTLESTONE, Jess *et al.* Artificial Intelligence in a Crisis Needs Ethics With Urgency. *Nature Machine Intelligence*, 22 jun. 2020. Disponível em: nature.com/articles/s42256-020-0195-0. Acesso em: 14 mar. 2022.

Disponível em: <https://bit.ly/3r4DXLZ>. Acesso em: 01 abr. 2022.

ETHICS and Governance of ai at Berkman Klein: Report on Impact, 2017-2019. Berkman Klein Center, 11 out. 2019. Disponível em: <https://bit.ly/3DJrsdJ>. Acesso em: 14 mar. 2022.

Ver TRICKEY, Erick. The Law and the Digital World. *Harvard Law Today*, 3 abr. 2019. Disponível em: <https://bit.ly/3u814Hs>. Acesso em: 14 mar. 2022.

HASSE, Alexa *et al.* Youth and Artificial Intelligence: Where We Stand. *Berkman Klein Center for Internet and Society*, 2019. Disponível em: <https://bit.ly/37j39ao>. Acesso em: 14 mar. 2022.

Relatório do *workshop* disponível em: ai4children.splashthat.com. Acesso em: 14 mar. 2022.

Catálogo disponível em: <https://bit.ly/3jq28At>. Acesso em: 14 mar. 2022.

FLORIDI, Luciano; COWLS, Josh. A Unified Framework of Five Principles for ai in Society. *Harvard Data Science Review*, 01 jul. 2019. Disponível em: <https://bit.ly/3uPhsvC>. Acesso em: 14 mar. 2022.

HLEGAI [High Level Expert Group on Artificial Intelligence]. Ethics Guidelines for Trustworthy AI. *European Commission*, 18 abr. 2019. Disponível em: <https://bit.ly/3r4DXLZ>. Acesso em: 14 de mar. 2022.

OECD. Recommendation of the Council on Artificial Intelligence. *OECD Legal Instruments*, 21 maio 2019. Disponível em: <https://bit.ly/3r6a3H5>. Acesso em 19 abr. 2022.

Disponível em: <https://bit.ly/3KfHtKS>. Acesso em: 1 abr. 2022.

O'Neil, Cathy. (2016). *Algoritmos de destruição em massa: como o Big Data aumenta a desigualdade e ameaça a democracia*. Trad. Rafael Abraham. Santo André: Rua do Sabão, 2020.

Ver “Surpreendentemente, Londres tem mais câmeras de vigilância do que Pequim”, coluna da *Época Negócios*, publicada em 19 de fevereiro de 2021, que compõe esta coletânea.

MÖKANDER, Jakob; FLORIDI, Luciano. Ethics-Based Auditing to Develop Trustworthy ai. *Minds and Machines*, v. 31, 19 fev. 2021. Disponível em: <https://bit.ly/36YnT7x>. Acesso em: 14 mar. 2022.

FIRST Draft of Recommendation on Ethics of Artificial Intelligence. Unesco, 7 set. 2020. Disponível em: <https://bit.ly/3x8l4LY>. Acesso em: 14 mar. 2022.

LEMOS, Ronaldo. Estratégia de ia brasileira é patética. *Folha de S.Paulo*, 11 abr. 2021. Disponível em: <https://bit.ly/3NQc43Z>. Acesso em: 14 mar. 2022.

MORIN, Edgar. Um festival de incerteza. Trad. Edgar Carvalho e Fagner França. *Instituto Humanitas Unisinos*, 9 jun. 2020. Disponível em: <https://bit.ly/3jdeXO4>. Acesso em: 14 mar. 2022.

GROTHER, Patrick; NGAN, Mei; HANAOKA, Kayee. Face Recognition Vendor Test (frvt). *National Institute of Standards and Technology, U. S. Department of Commerce*, [s.d.]. Disponível em: doi.org/10.6028/NIST.IR.8280. Acesso em: 14 mar. 2022.

Ver “Sistemas de vigilância com reconhecimento facial: escrutínio e reação das gigantes de tecnologia”, coluna da *Época Negócios*, publicada em 19 de junho de 2020, que compõe esta coletânea.

JOHNSON, Khari. AI Weekly: A Deep Learning Pioneer’s Teachable Moment on AI Bias. *VentureBeat*, 26 jun. 2020. Disponível em: <https://bit.ly/3jdfDmA>. Acesso em: 14 mar. 2022.

Idem.

BISCHOFF, Paul. Surveillance Camera Statistics: Which Cities Have the Most cctv Camera? *Comparitech*, [s.d.]. Disponível em: <https://bit.ly/3x7cVax>. Acesso em: 14 mar. 2022.

Ver “Sistemas de vigilância digital: combatem o coronavírus, mas ameaçam a privacidade e a liberdade”, coluna da *Época Negócios*, publicada em 27 de março de 2020, que compõe esta coletânea.

CRiado PEREZ, Caroline. *Invisible Women: Exposing Data Bias in a World Designed for Men*. London: Vintage, 2021.

WEST, Mark; KRAUT, Rebecca; CHEW, Han Ei. I’d Blush, If I Could: Closing Gender Divides in Digital Skills Through Education. *Equals*, Unesco, 2019. Disponível em: <https://en.unesco.org/ld-blush-if-i-could>. Acesso em: 14 mar. 2022.

Disponível em: <https://nyti.ms/37vVoyj>. Acesso em: 14 mar. 2022.

Disponível em: <https://bit.ly/3rEF6KB>. Acesso em: 14 mar. 2022.

Disponível em: <https://bit.ly/3KZYF7A>. Acesso em: 14 mar. 2022.

Disponível em: <https://bit.ly/392PFk9>. Acesso em: 14 mar. 2022.

Disponível em: <http://glo.bo/3vwLZi2>. Acesso em: 14 mar. 2022.

HARWELL, Drew. The Accent Gap. *The Washington Post*, 19 jul. 2018. Disponível em: <https://wapo.st/3Jcg2R3>. Acesso em: 14 mar. 2022.

KOENECKE, Allison *et al.* Racial Disparities in Automated Speech Recognition. *pnas*, 7 abr. 2020. Disponível em: pnas.org/content/117/14/7684. Acesso em: 14 mar. 2022.

LIMA, Lanna *et al.* Empirical Analysis of Bias in Voice-based Personal Assistants. In: *www’19: Companion Proceedings of the 2019 World Wide Web Conference*, New York, maio 2019. Disponível em: <https://>

- bit.ly/3j6zjZl. Acesso em: 14 mar. 2022.
- LESLIE, David. Understanding Bias in Facial Recognition Technologies. *The Alan Turing Institute*, 26 set. 2020. Disponível em: <https://bit.ly/36Yeftt>. Acesso em: 14 mar. 2022.
- CHRISTIAN, Brian. *The Alignment Problem: Machine Learning and Human Values*. New York: W. W. Norton & Company, 2020.
- DUHIGG, Charles. How Companies Learn Your Secrets. *The New York Times*, 16 fev. 2012. Disponível em: <https://nyti.ms/3Kg5GRn>. Acesso em: 14 mar. 2022; *O poder do hábito*. Trad. Rafael Mantovani. Rio de Janeiro: Objetiva, 2012.
- FRY, Hannah. *Hello Word: Being Human in the Age of Algorithms*. New York: W. W. Norton & Company, 2018.
- BENKLER, Yochai; FARIS, Robert; ROBERTS, Hal. *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. New York: Oxford University Press, 2018.
- SIMONDON, Gilbert. (1958). *Do modo de existência dos objetos técnicos*. Trad. Vera Ribeiro. Rio de Janeiro: Contraponto, 2020.
- ZUBOFF. *A era do capitalismo de vigilância*.
- FOGG, Brian. Persuasive computers: perspectives and research directions. *Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY: Addison-Wesley Publishing Co., 1998. p. 225-232. Disponível em: <https://bit.ly/3DGjeTX>. Acesso em: 2 abr. 2022.
- SHIRKY, Clay. (2010). *A cultura da participação: criatividade e generosidade no mundo conectado*. Trad. Celina Portocarrero. Rio de Janeiro: Zahar, 2011.
- FERNANDES, Daniela. “Gigantes da web são novo Estado”, diz Pierre Lévy. *Valor Econômico*, 23 out. 2020. Disponível em: <http://glo.bo/35ITFVz>. Acesso em: 14 mar. 2022.
- SIMONS, Josh; GHOSH, Dipayan. Utilities for Democracy: Why and How the Algorithmic Infrastructure of Facebook and Google Must Be Regulated. *Brookings*, ago. 2020. Disponível em: <https://brook.gs/3JbXkZG>. Acesso em: 14 mar. 2022.
- WU, Tim. *The Curse of Bigness: Antitrust in the New Gilded Age*. New York: Columbia Global Reports, 2018.
- KIRKPATRICK, David. The Facebook Defect. *Time*, 12 abr. 2018. Disponível em: <https://time.com/5237458/the-facebook-defect/>. Acesso em: 14 mar. 2022.
- AL-SALEH, Asaad. *Voices of the Arab Spring*. New York: Columbia University Press, 2015.
- FILGUEIRAS, Fernando; ALMEIDA, Virgílio. *Governance for the Digital World: Neither More State nor More Market*. London: Palgrave Macmillan, 2021.
- HUSSAIN, Waheed. The Common Good. In: STANFORD Encyclopedia of Philosophy. Disponível em: <https://stanford.io/3KdRaK1>. Acesso em: 14 mar. 2022.
- BENKLER, Yochai. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven: Yale University Press, 2006.

BENKLER, Yochai. Law, Innovation, and Collaboration in Networked Economy and Society. *Annual Review of Law and Social Science*, v. 13, 2017. Disponível em: <https://bit.ly/3KgeMxt>. Acesso em: 14 mar. 2022.

Disponível em: <https://tcrn.ch/38wMf95>. Acesso em: 02 abr. 2022.

HARARI. *Homo Deus*.

HALO Privacy Whitepaper. Disponível em: <https://bit.ly/36QByxK>. Acesso em: 14 mar. 2022.

FRISCHMANN, Brett; SELINGER, Evan. *Re-Engineering Humanity*. Cambridge: Cambridge University Press, 2018.

KHAN, Lina M. *Amazon's Antitrust Paradox*. *Yale Law Journal*, v. 126, 2016. Disponível em: <https://bit.ly/372cVhk>. Acesso em: 14 mar. 2022.

WU, Tim. *The Curse of Bigness: Antitrust in the New Gilded Age*. New York: Columbia Global Report, 2018.

Disponível em: <https://on.wsj.com/3rBQTsO>. Acesso em: 14 mar. 2022.

Ver “Proposta europeia de regulamentação da inteligência artificial: impressões preliminares”, coluna da *Época Negócios*, publicada em 30 de abril de 2021, que compõe esta coletânea.

MAYER-SCHÖNBERGER, Viktor; RAMGE, Thomas. *Reinventing Capitalism in the Age of Big Data*. New York: Basic Book, 2018.

Disponível em: <https://bit.ly/3NLoFWl>. Acesso em: 14 mar. 2022.

Ver “Lina Khan, presidente da FTC: ameaça vigorosa ao poder das gigantes de tecnologia”, coluna da *Época Negócios*, publicada em 25 de junho de 2021, em compõe esta coletânea.

SEN, Amartya. *Sobre ética e economia*. Trad. Laura Teixeira Motta. São Paulo: Companhia das Letras, 1999.

Disponível em: <https://bit.ly/3J8vj5f>. Acesso em: 14 mar. 2022.

MAYER-SCHÖNBERGER; RAMGE. *Reinventing Capitalism in the Age of Big Data*. NY: Basic Books, 2018.

FRISCHMANN; SELINGER. *Re-Engineering Humanity*. UK: Cambridge University Press, 2018.

VELASQUEZ-MANOFF, Moises. The Brains Implant That Could Change Humanity. *The New York Times*, 28 ago. 2020. Disponível em: <https://nyti.ms/3NOUJZe>. Acesso em: 14 mar. 2022.

BOSTROM, Nick. (2014). *Superinteligência: caminhos, perigos e estratégias para um novo mundo*. Trad. Aurélio Antônio Monteiro et al. Rio de Janeiro: Darkside, 2018.

Entrevista realizada no ciclo de debates Conversações, do tiddigital, promovido pelo Programa de Estudos Pós-Graduados em Tecnologias da Inteligência e Design Digital da puc-sp, em 5 de agosto de 2020. Disponível em: <https://youtu.be/0PrG5IHBLVk>. Acesso em: 14 mar. 2022.

LANDHUIS, Esther. Neuroscience: Big Brain, Big Data. *Nature*, v. 541, p. 559-561, 26 jan. 2017. Disponível em: doi.org/10.1038/541559a. Acesso em: 14 mar. 2022.

BORA, Adriana; TIMIS, David Alexandru. We Need to Talk About Artificial Intelligence. *The World Economic Forum*, 22 fev. 2021. Disponível em: <https://bit.ly/3LMk587>. Acesso em: 14 mar. 2022.

PROPOSAL for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. *Comissão Europeia*, Bruxelas, 21 abr. 2021. Disponível em: <https://bit.ly/3DGFT2q>. Acesso em: 14 mar. 2022.

SALOMÃO, Luís Felipe (Org.). *Inteligência artificial: tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário Brasileiro*. fgv Conhecimento; Centro de Inovação, Administração e Pesquisa do Judiciário, 2020. Disponível em: <https://bit.ly/3DHmpdX>. Acesso em: 14 mar. 2022.

MAIA FILHO, Mamede Said; JUNQUILHO, Tainá Aguiar. Projeto Victor: perspectivas de aplicação da inteligência artificial ao direito. *Revista de Direitos e Garantias Fundamentais*, v. 19, n. 3, 2018. Disponível em: <https://bit.ly/3LGgqcb>. Acesso em: 14 mar. 2022.

CONSELHO NACIONAL DE JUSTIÇA. Resolução n.º 332, de 31 de agosto de 2020. Dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências. Brasília, 2020. Disponível em: <https://bit.ly/3KcTm4z>. Acesso em: 14 mar. 2022.

Asilomar AI Principles. Disponível em: futureoflife.org/ai-principles. Acesso em: 14 mar. 2022.

PROPOSAL for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. *Comissão Europeia*, Bruxelas, 21 abr. 2021. Disponível em: <https://bit.ly/3DGFT2q>. Acesso em: 14 mar. 2022.

DOCHERTY, Neil; FANNING, David. [S. l.: s. n.], 2019. 1 vídeo (1h 54min 16s). In the Age of AI. Publicado pelo canal Frontline PBS. Disponível em: https://youtu.be/5dZ_lvDgevk. Acesso em: 14 mar. 2022..

DOCHERTY, Neil; FANNING, David. [S. l.: s. n.], 2019. 1 vídeo (1h 54min 16s). In the Age of AI. Publicaod pelo canal Frontline PBS. Disponível em: https://youtu.be/5dZ_lvDgevk. Acesso em: 14 mar. 2022.

ROBERTS, Huw; COWLS, Josh; MORLEY, Jessica *et al.* The Chinese Approach to Artificial Intelligence: An Analysis of Policy, Ethics, and Regulation. *AI & Society*, v. 36, 2021. Disponível em: <https://bit.ly/35IW16R>. Acesso em: 14 mar. 2022.

O'MEARA, Sarah. AI Researchers in China Want to Keep the Global-sharing Culture Alive. *Nature*, v. 569, 29 maio 2019. Disponível em: <https://go.nature.com/3M0OrnR>. Acesso em: 14 mar. 2022.

THE AI Index Report: Measuring Trends in Artificial Intelligence. *Stanford Institute for Human-Centered Artificial Intelligence*, 2021. Disponível em: <https://stanford.io/3Kg8C0l>. Acesso em: 14 mar. 2022.

Ver “Proposta europeia de regulamentação da inteligência artificial: impressões preliminares”, coluna da *Época Negócios*, publicada em 30 de abril de 2021, que compõe esta coletânea.

FLORIDI, Luciano. AI and Its New Winter: from Myths to Realities. *Philosophy and Technology*, v. 33, 2020. Disponível em: <https://bit.ly/3j65RD1>. Acesso em: 14 mar. 2022.

MINSKY, Marvin; PAPERT, Seymour A. *Perceptrons: An Introduction to Computational Geometry*. Cambridge: MIT Press, 1969.

Ver “Inteligência artificial e as idiossincrasias do poder público brasileiro”, coluna da *Época Negócios*, publicada em 9 de julho de 2021, que compõe esta coletânea.

FLORIDI, Luciano. The End of an Era: from Self-Regulation to Hard Law for the Digital Industry. *Philosophy and Technology*, v. 34, 2021. Disponível em: <https://bit.ly/3r4gg6G>. Acesso em: 14 mar. 2022.

Disponíveis em: blog.google/technology/ai/ai-principles e pair.withgoogle.com/guidebook. Acesso em: 14 mar. 2022.

JOHNSON, Bobbie; LICHFIELD, Gideon. Hey Google, Sorry You Lost Your Ethics Council, so We Made One for You. *MIT Technology Review*, 6 abr. 2019. Disponível em: <https://bit.ly/3NNbbcz>. Acesso em: 14 mar. 2022.

Ver “Proposta europeia de regulamentação da inteligência artificial: impressões preliminares”, coluna da *Época Negócios*, publicada em 30 de abril de 2021, que compõe esta coletânea.

Ver KAUFMAN, Dora. Um projeto de futuro. *Piauí*, 22 out. 2021. Disponível em: <https://bit.ly/36SUw6S>. Acesso em: 14 mar. 2022.

BOSTON UNIVERSITY SCHOOL OF MEDICINE. Researchers Enhance Alzheimer’s Disease Classification through Artificial Intelligence. *Medical Xpress*, 15 mar. 2021. Disponível em: <https://bit.ly/3DIkwh2>. Acesso em: 14 mar. 2022.

TOPOL, Eric. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. New York: Basic Books, 2019.

ADLER, Samuel; METZGER, Yoav. PillCam Colon Capsule Endoscopy: Recent Advances and New Insights. *Therapeutic Advances in Gastroenterology*, v. 4, n. 4, jul. 2011. Disponível em: <https://bit.ly/3j4EKs1>. Acesso em: 14 mar. 2022.

MATWYSHYN, Andrea M. The “Internet of Bodies” is Here. Are Courts and Regulators Ready? *The Wall Street Journal*, 12 nov. 2018. Disponível em: <https://on.wsj.com/36Yt739>. Acesso em: 14 mar. 2022.

LIU, Xiao; MERRITT, Jeff. Shaping the Future of the Internet of Bodies: New Challenges of Technology Governance. Briefing Paper. *The World Economic Forum*, jul. 2020. Disponível em: <https://bit.ly/3ukWUMF>. Acesso em: 14 mar. 2022.

ISHIGURO, Kazuo. *Klara e o Sol*. Trad. Ana Guadalupe. São Paulo: Companhia das Letras, 2021.

DAVIS, Nicola. Scientists Create Online Games to Show Risks of AI Emotion Recognition. *The Guardian*, 4 abr. 2021. Disponível em: <https://bit.ly/3xALkPm>. Acesso em: 14 mar. 2022.

Woebot Health. Disponível em: <https://woebothealth.com/about-us>. Acesso em: 2 abr. 2022.

DARCY, Alison *et al.* Evidence of Human-Level Bonds Established With a Digital Conversational Agent: Cross-sectional, Retrospective Observational Study. *JMIR Formative Research*, v. 5, n. 5, 11 maio 2021. Disponível em: <https://bit.ly/3JYzVeJ>. Acesso em: 14 mar. 2022.

KENT, Chloe. Mental Health Chatbots might do Better When they don’t Try to Act Human. *Medical Device Network*, 21 maio 2021. Disponível em: <https://bit.ly/3LI2DSj>. Acesso em: 2 abr. 2022..

ABOUJAOUDE, Elias. Digital Interventions in Mental Health: Current Status and Future Directions.

- Frontiers in Psychiatry*, v. 11, fev. 2020. Disponível em: <https://bit.ly/3j8Ns8z>. Acesso em: 14 mar. 2022.
- LEMBKE, Anna. *Nação dopamina: por que o excesso de prazer está nos deixando infelizes e o que podemos fazer para mudar*. São Paulo: Vestígio, 2022.
- Disponível em: aspgh.org/opioids. Acesso em: 14 mar. 2022.
- U.S. OPIOID Dispensing Rate Maps. Disponível em: cdc.gov/drugoverdose/rxrate-maps/index.html. Acesso em: 14 mar. 2022.
- CASE, Anne; DEATON, Angus. *Deaths of Despair and the Future of Capitalism*. Princeton; Oxford: Princeton University Press, 2020.
- BOMBANA, Henrique Silva *et al.* Use of Alcohol and Illicit Drugs by Trauma Patients in Sao Paulo, Brazil. *Injury – International Journal of the Care of the Injured*, 30 out. 2021. Disponível em: <https://bit.ly/3NSwkSI>. Acesso em: 14 mar. 2022.
- Ver “Nem tudo é tecnologia e seus efeitos: é preciso identificar as causas dos problemas sociais”, coluna *Época Negócios*, publicada em 7 de janeiro de 2022, que compõe esta coletânea.
- FORTIM, Ivelise; SPRITZER, Daniel Tornaim; LIMA, Maria Thereza Alencar. *Games viciam: fato ou ficção?*. São Paulo: Estação das Letras e Cores, 2020.
- OXFAM. Inequality Kills. *Oxfam*, 16, jan. 2022. Disponível em: <https://bit.ly/3JaRSXa>. Acesso em: 14 mar. 2022.
- MOROZOV, Evgeny. *Big tech: a ascensão dos dados e a morte da política*. Trad. Claudio Marcondes. São Paulo: Ubu, 2018.
- WHO. Report of the WHO-China Joint Mission on Coronavirus Disease. 16-24 fev. 2020. Disponível em: <https://bit.ly/3DQMuHG>. Acesso em: 14 mar. 2022.
- ARTIFICIAL Intelligence for COVID-19: Saviour or Saboteur? *The Lancet Digital Health*, 22 dez. 2020. Disponível em: <https://bit.ly/36QNb0l>. Acesso em: 14 mar. 2022.
- HASHIGUCHI, Tiago Cravo Oliveira. Artificial intelligence in health still needs human intelligence: The COVID-19 crisis has shown that AI can deliver benefits, but it has also exposed its limits, often related to having the right data. *Health Policy Analyst*, OECD, 15 set. 2021. Disponível em: <https://bit.ly/3jU38gi>. Acesso em: 19 abr. 2022.
- HAN, Byung-Chul. O coronavírus de hoje e o mundo de amanhã. *El País*, 22 mar. 2020. Disponível em: <https://bit.ly/3ud57IZ>. Acesso em: 14 mar. 2022.
- MOZUR, Paul; ZHONG, Raymond; KROLIK, Aaron. In Coronavirus Fight, China Gives Citizens a Color Code, With Red Flags. *The New York Times*, 01 mar. 2020, disponível em: <https://nyti.ms/3DMGvDM>. Acesso em: 14 mar. 2022.
- KHARPAL, Arjun. Coronavirus could be a “catalyst” for China to boots its mass surveillance machine, experts say. *CNBC*, 24 fev. 2020. Disponível em: <https://cnb.cx/3j4EDg2>. Acesso em: 2 abr. 2022.
- JOHNSON, Khari. How People are Using AI to Detect and Fight the Coronavirus. *VentureBeat*, 3 mar. 2020. Disponível em: <https://bit.ly/3Jdm97q>. Acesso em: 14 mar. 2022.

JOHNSON, Khari. DarwinAI Wants to Help Identify Coronavirus in X-rays, but Radiologists Aren't Convinced. *VentureBeat*, 25 mar. 2020. Disponível em: <https://bit.ly/3r5wLiK>. Acesso em: 14 mar. 2022.

RESEARCHERS Will Deploy AI to Better Understand Coronavirus. *Wired*, 17 mar. 2020. Disponível em: <https://bit.ly/3713DSN>. Acesso em: 14 mar. 2022.

WYNANTS, Laure *et al.* Prediction Models for Diagnosis and Prognosis of Covid-19 Infection: Systematic Review and Critical Appraisal. *The British Medical Journal*, 7 abr. 2020. Disponível em: <https://bit.ly/36Ylbiv>. Acesso em: 14 mar. 2022.

FORSTER, E. M. *A máquina parou*. Org. e trad. Teixeira Coelho. São Paulo: Itaú Cultural; Iluminuras, 2018.

CARBAJOSA, Ana. Peter Sloterdijk: “O regresso à frivolidade não vai ser fácil”. Entrevista. *El País*, 9 maio 2020. Disponível em: <https://bit.ly/3ubA0al>. Acesso em: 14 mar. 2022.

MELLO, Michelle M.; WANG, C. Jason. Ethics and Governance for Digital Disease Surveillance. *Science*, 29 maio 2020. Disponível em: <https://bit.ly/3DFVDMh>. Acesso em: 14 mar. 2022.

ANGWIN, Julia; LARSON, Jeff; MATTU, Surya; KIRCHNER, Lauren. Machine Bias. *ProPublica*, 23 maio 2016. Disponível em: <https://bit.ly/3KfBi9U>. Acesso em: 14 mar. 2022.

BUOLAMWINI, Joy; GEBRU, Timnit. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, v. 81, p. 1-15, 2018. Disponível em: <https://bit.ly/3j7iA8j>. Acesso em: 14 mar. 2022.

SILBERG, Jake; MANYIKA, James. Tackling Bias in Artificial Intelligence (and in Humans). *McKinsey Global Institute*, 6 jun. 2019. Disponível em: <https://bit.ly/36SZXmi>. Acesso em: 14 mar. 2022.

Disponível em: <https://bit.ly/3Ke3gCV>. Acesso em: 14 mar. 2022.

TOEWS, Rob. Deep Learning's Carbon Emissions Problem. *Forbes*, 17 jun. 2020. Disponível em: <https://bit.ly/3j8EOag>. Acesso em: 14 mar. 2022.

ANDREWS, Edmund L. AI's Carbon Footprint Problem. *Stanford University Human-Centered Artificial Intelligence Blog*, 2 jul. 2020. Disponível em: <https://stanford.io/38wfnge>. Acesso em: 14 mar. 2022.

ANDREWS, Edmund L.

UnHerd (revista online), “How AI will terminate us”, Peter Franklin, 11 jun. 2019. Disponível em: <https://bit.ly/3LHTVUi>. Acesso em: 2 abr. 2022.

ROLNICK, David *et al.* Tackling Climate Change with Machine Learning. *ArXiv*, 5 nov. 2019. Disponível em: arxiv.org/pdf/1906.05433.pdf. Acesso em: 14 mar. 2022.

CRAWFORD, Kate. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven; London: Yale University Press, 2021.

STRUBELL, Emma; GANESH, Ananya; MCCALLUM, Andrew. Energy and Policy Considerations for Deep Learning in NLP. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, jan. 2019. Disponível em: <https://bit.ly/3v13wil>. Acesso em: 14 mar. 2022.

COOK, Gary *et al.* Clicking Clean: Who is Winning the Race to Build a Green Internet? *Greenpeace*, jan. 2017. Disponível em: <https://bit.ly/3uaYVdU>. Acesso em: 14 mar. 2022.

POWERING the Cloud: How China's Internet Industry Can Shift to Renewable Energy. *Greenpeace*, 2019. Disponível em: <https://bit.ly/3uYrBGd>. Acesso em: 14 mar. 2022.

Disponível em: <https://bit.ly/35HtSNt>. Acesso em: 14 mar. 2022.

Disponível em: <https://bit.ly/3DQTfjy>. Acesso em: 14 mar. 2022.

Disponível em: <https://bit.ly/3LAdbmy>. Acesso em: 14 mar. 2022.

Disponível em: <https://bit.ly/3uazklo>. Acesso em: 2 abr. 2022.

CLIMA e Desenvolvimento: Visões para o Brasil 2030. *iCS - Clima e Sociedade*, 10 nov. 2021. Disponível em: <https://bit.ly/3uazklo>. Acesso em 19 abr. 2022.

VINUESA, R.; AZIZPOUR, H.; LEITE, I. *et al.* The Role of Artificial Intelligence in Achieving the Sustainable Development Goals. *Nature Communications*, 2020. Disponível em: <https://go.nature.com/3uVZ30e>. Acesso em: 14 mar. 2022.

Disponível em: <https://bit.ly/37pe4Qc>. Acesso em: 2 abr. 2022.

VARGA, George. How Peter Jackson shifted the dour context of “Let It Be” into a fuller, more human “Get Back”. *The Washington Post*, 25 nov. 2021. Disponível em: <https://bit.ly/3uv2lZq>. Acesso em: 11 abr. 2022.

BEETHOVEN: Exclusive insight into the 10th symphony. [S. l.: s. n.], 20 set. 2021. 1 vídeo (4min 10s). Publicado pelo canal Deutsche Telekom. Disponível em: <https://bit.ly/3xhK7w2>. Acesso em: 11 abr. 2022.

AI & ARTS: How can AI augment and be enriched by the arts and how far can data science and the arts help to answer each other's questions?. *The Alan Turing Institute*. Disponível em: <https://bit.ly/3O4AW8r>. Acesso em: 11 abr. 2022.

CHALMERS, David J. *Reality+: Virtual Worlds and the Problems of Philosophy*. Londres: Penguin, 2022.

2º CONGRESSO de Inteligência Artificial da PUC-SP – Dia 2. [S. l.: s. n.], 18 nov. 2021. 1 vídeo (7h 47min 41s). Publicado pelo canal TIDD PUC-SP. Disponível em: <https://bit.ly/3KySszg>. Acesso em: 11 abr. 2022.

CONVERSACÕES – Dora Kaufman e Roberto Lent – Redes neurais cérebro humano. [S. l.: s. n.], 5 ago. 2020. 1 vídeo (1h 03min 32s). Publicado pelo canal TIDD PUC-SP. Disponível em: <https://bit.ly/3v4nKaO>. Acesso em: 11 abr. 2022.

SMITH, Brad; SHUM, Harry. Foreword: The Future Computed. In: *The Future Computed: Artificial Intelligence and its Role in Society*. Washington: Microsoft, 2018.

WILDE, Oscar. *The Decay of Lying and Other Essays*. London: Penguin Books, 2020.

Edição brasileira: *A Inteligência Artificial a Nosso Favor: como manter o controle sobre a tecnologia*, Companhia das Letras, 2021.

SCHNEIDER, Susan. *Artificial You: AI and The Future of Your Mind*. Princeton: Princeton University Press, 2019.

TURKLE, Sherry. *Alone Together: Why We Expect More from Technology and Less from Each Other*. New York: Basic Books, 2017

EZRACHI, Ariel; STUCKE, Maurice. *Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy*. Cambridge: Harvard University Press, 2016.

HAN, Byung-Chul. (2012). *A sociedade da transparência*. Trad. Enio Paulo Giachini. Petrópolis: Vozes, 2017.

ALBERTI, Fay Bound. *A Biography of Loneliness: The History of an Emotion*. New York: Oxford University Press, 2019.

Disponível em: <https://bit.ly/36YqUF5>. Acesso em: 2 abr. 2022.

ROSS, Alec. *The Industries of the Future*. New York: Simon & Schuster, 2016.

BOSTROM, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2014.

A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, 1956. Disponível em: <https://bit.ly/3xIKNea>. Acesso em: 15 abr. 2022.

RUSSELL, Stuart. *Human Compatible: Artificial Intelligence and the Problem of Control*. New York: Viking; Penguin, 2019.

MITCHELL, Melanie. Why AI is Harder Than we Think. *ArXiv*, 28 abr. 2021. Disponível em: arxiv.org/abs/2104.12871. Acesso em: 14 mar. 2022.

GUSTIN, Sam. IBM'S Watson Supercomputer Wins Practice Jeopardy Round. *Wired*, 13 jan. 2011. Disponível em: <https://bit.ly/37jcWNE>. Acesso em: 14 mar. 2022.

SHEAD, Sam. Google DeepMind is Edging Towards a 3-0 Victory against World Go Champion Ke Jie. *Insider*, 25 maio 2017. Disponível em: <https://bit.ly/3DIyVK8>. Acesso em: 14 mar. 2022.

McDERMOTT, Drew. Artificial Intelligence Meets Natural Stupidity. *ACM SIGART Bulletin*, v. 57, abr. 1976. Disponível em: <https://bit.ly/38zQoct>. Acesso em: 14 mar. 2022.

CHEN, Alicia; LI, Lyric. China's Lonely Hearts Reboot Online Romance with Artificial Intelligence. *The Washington Post*, 6 ago. 2021. Disponível em: <https://wapo.st/35LSOUc>. Acesso em: 14 mar. 2022.

LOVELOCK, James. *Novacene: The Coming Age of Hyperintelligence*. London: Penguin, 2019.

SKJUVE, Marita *et al.* My Chatbot Companion: A Study of Human-Chatbot Relationship. *International Journal of Human-Computer Studies*, v. 149, maio 2021. Disponível em: <https://bit.ly/3NQgsjG>. Acesso em: 14 mar. 2022.

KALAJIAN, Rob. How Many People Play Fortnite in 2021?. *Sportskeeda*, 27 jan. 2021. Disponível em: <https://bit.ly/3OnXgd1>. Acesso em: 14 mar. 2022.

STEPHENSON, Neal. *Snow Crash*. Trad. Fábio Fernandes. São Paulo: Aleph, 2015.

- HODENT, Celia. Understanding the Success of Fortnite: A UX & Psychology Perspective. *Brains, UX & Games*, 22 jul. 2019. Disponível em: <https://bit.ly/3DXtzuV>. Acesso em: 14 mar. 2022; *The Gamer's Brain: How Neuroscience and UX Can Impact Video Game Design*. Boca Raton: Taylor & Francis, 2018.
- WU, Tim. *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*. New York: Knopf, 2016.
- Afirmção relativamente conhecida, replicada em inúmeras fontes. Disponível em: <https://bit.ly/3uSpwwN>. Acesso em: 17 abr. 2022.
- Disponível em: <https://bbc.in/3EtKg18>. Acesso em: 17 abr. 2022.
- Disponível em: <https://bit.ly/3JY2bhJ>. Acesso em: 17 abr. 2022.
- BOBBY Fischer, génie paranoïaque des échecs. *Le Monde*, 19 jan. 2008. Disponível em: <https://bit.ly/3LDUSga>. Acesso em: 14 mar. 2022.
- ALBERGOTTI, Reed. He Predicted the Dark Side of the Internet 30 Years Ago. Why Did No One listen? *The Washington Post*, 12 ago. 2021. Disponível em: <https://wapo.st/36QPRCs>. Acesso em: 14 mar. 2022.
- AGRE, Philip E. Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI. In: BOWKER, Geoff; GASSER, Les; STAR, Leigh; TURNER, Bill (Eds.). *Bridging the Great Divide: Social Science, Technical Systems, and Cooperative Work*. Los Angeles: Erlbaum, 1997. Disponível em: <https://bit.ly/3uZunLo>. Acesso em: 2 abr. 2022.
- BURKE, Peter. *O polímata: uma história cultural de Leonardo da Vinci a Susan Sontag*. Trad. Renato Prelorenzou. São Paulo: Editora Unesp, 2020.
- LEE, Kai-Fu; QIU-FAN, Chen. *AI 2041: Ten Visions of Our Future*. New York: Currency, 2021.
- SHLADOVER, Steven. "Self-Driving" Cars Begin to Emerge from a Cloud of Hype. *Scientific American*, 25 set. 2021. Disponível em: <https://bit.ly/3uc2YXF>. Acesso em: 14 mar. 2022.
- SANTAELLA, Lucia. *Humanos hiper-híbridos: linguagens e cultura na segunda era da internet*. São Paulo: Paulus, 2021.
- FFHC Debate: capitalismo de vigilância e democracia, com Shoshana Zuboff. *Fundação FHC*, 9 dez. 2021. Disponível em: <https://bit.ly/3r7oHy2>. Acesso em: 14 mar. 2022.
- KROGSTAD, Jens Manuel; LOPEZ, Mark Hugo. Black Voter Turnout Fell in 2016, even as a Record Number of Americans Cast Ballots. *Pew Research Center*, 12 maio 2017. Disponível em: <https://pewrs.ch/37mcq1G>. Acesso em: 14 mar. 2022.
- Ver: BANCO CENTRAL DO BRASIL. Relatório anual 2010. *Boletim do Banco Central do Brasil*, 2010. Disponível em: <https://bit.ly/3LOJoqt>; Relatório anual 2015. *Boletim do Banco Central do Brasil*, 2015. Disponível em: <https://bit.ly/36T69e2>; Estatísticas monetárias e de crédito. *Banco Central do Brasil*, jan. 2022. Disponível em: <https://bit.ly/3jbnDVg>. Acesso em: 14 mar. 2022.
- DAVIS, Kenrick. Welcome to China's latest "robot restaurant". *The World Economic Forum*, 01 jul. 2020. Disponível em: <https://bit.ly/361IQPF>. Acesso em: 11 abr. 2022.
- GATHERING Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report. *Stanford University*, set. 2021. Disponível em: <https://stanford.io/3LXzvXu>. Acesso em: 11 abr. 2022.

ZHANG, Baobao; DAFOE, Allan. Artificial Intelligence: American Attitudes and Trends. *Social Science Research Network*, 19 jan. 2019. Disponível em: <https://bit.ly/3xknjf5>. Acesso em: 11 abr. 2022.

SHERMAN, Jason. Russia-Ukraine Conflict Prompted U.S. to Develop Autonomous Drone Swarms, 1.000-Mile Cannon. *Scientific American*, 14 fev. 2022. Disponível em: <https://bit.ly/3uxfuRq>. Acesso em: 11 abr. 2022.

KALLENBORN, Zachary. Russia may have used a killer robot in Ukraine. Now what?. *Bulletin of the Atomic Scientists*, 15 mar. 2022. Disponível em: <https://bit.ly/3v4XME6>. Acesso em: 11 abr. 2022.

KNIGHT, Will. Russia's Killer Drone in Ukraine Raises Fears About AI in Warfare. *Wired*, 17 mar. 2022. Disponível em: <https://bit.ly/3O3huJ6>. Acesso em: 11 abr. 2022.

CHAMAYOU, Grégoire. *Teoria do drone*. São Paulo: Cosac & Naify, 2015.

DE VYNCK, Gerrit. The U.S. Says Humans Will Always Be in Control of AI Weapons: But the Age Of Autonomous War is Already Here. *The Washington Post*, 7 jul. 2021. Disponível em: <https://wapo.st/3ju2mGt>. Acesso em: 11 abr. 2022.

Ibidem.

SHERMAN, Jason. Russia-Ukraine Conflict Prompted U.S. to Develop Autonomous Drone Swarms, 1.000-Mile Cannon. *Scientific American*, 14 fev. 2022. Disponível em: <https://bit.ly/3uxfuRq>. Acesso em: 11 abr. 2022.

KALLENBORN, Zachary. Russia may have used a killer robot in Ukraine. Now what?. *Bulletin of the Atomic Scientists*, 15 mar. 2022. Disponível em: <https://bit.ly/3v4XME6>. Acesso em: 11 abr. 2022.

PhD no Physics Department and Artificial Intelligence Laboratory, do Massachusetts Institute of Technology (MIT), professor associado em Ciência da Computação no Courant Institute of Mathematical Sciences, New York University, e cofundador de *startups* de inteligência artificial.

Contents

1. Prefácio | Fábio Gagliardi Cozman
2. Introdução
3. Fundamentos e lógica da IA
 1. O que é a inteligência artificial hoje?
 2. A tecnologia avança aceleradamente, mas ainda distante da condição humana
 3. São as máquinas inteligentes?
 4. Redes neurais artificiais e a complexidade do cérebro humano
 5. Da personalização do discurso em Aristóteles à personalização com algoritmos de inteligência artificial
 6. Interpretabilidade e confiança: variáveis interdependentes nos modelos de redes neurais
 7. Base de dados para treinar algoritmo de IA não é salsicha: qualidade dos “ingredientes” é crítica
4. IA e mercado de trabalho
 1. Você está preparado para trabalhar no século XXI?
 2. Inteligência artificial pode democratizar o acesso à Justiça?
 3. Desigualdade crescente no mercado de trabalho
 4. Como evitar a classe dos “inempregáveis”?
 5. Inteligência artificial no jornalismo: ameaça ou oportunidade?
 6. A sustentabilidade do trabalho depende da coexistência positiva de humanos e IA
5. Justiça e ética na IA
 1. As fake news também atingem os negócios
 2. Os algoritmos de inteligência artificial podem ser éticos?
 3. Alerta: as tecnologias de reconhecimento facial estão nos ameaçando
 4. A objetividade relativa da IA pode neutralizar a subjetividade humana?
 5. Aplicativo de reconhecimento facial: como fica o direito ao anonimato?
 6. Ética é objeto da ação humana, não existe ética da inteligência artificial
 7. Princípios éticos gerais versus regras práticas: reflexões sobre o AI Guidebook do Google
 8. Documentário Coded Bias: o rosto como última fronteira da privacidade

6. A IA e o problema do viés
 1. Sistemas de vigilância com reconhecimento facial: escrutínio e reação das gigantes de tecnologia
 2. Ética by design no desenvolvimento de tecnologias
 3. Surpreendentemente, Londres tem mais câmeras de vigilância do que Pequim
 4. Lacuna de dados de gênero: mulheres invisíveis num mundo projetado por homens
 5. A interação humano-máquina por voz é potencialmente inclusiva, mas na prática tem sido discriminatória
7. A economia de dados e o poder das big techs
 1. Tecnologia gratuita em troca de seus dados: o melhor e o pior do capitalismo em um simples swap
 2. Benefícios e ameaças da automação inteligente do varejo
 3. Documentário O dilema das redes: a polêmica da vez
 4. “Estado-Plataforma”: o poder das big techs
 5. Dados gerados pelos usuários nas plataformas digitais: bens comuns ou proprietários?
 6. Tecnologias ambivalentes: conveniência versus pesadelo orwelliano
 7. Lina Khan, presidente da FTC: ameaça vigorosa ao poder das gigantes de tecnologia
 8. Desafio do capitalismo de dados: reaproximar economia e ética com base em novos paradigmas
8. Iniciativas regulatórias
 1. A sintonia do livro 1984 com as ameaças do século XXI
 2. Decodificando o cérebro humano: assimetria informacional entre produtores de tecnologia e legisladores
 3. Proposta europeia de regulamentação da inteligência artificial: impressões preliminares
 4. Inteligência artificial e as idiosincrasias do poder público brasileiro
 5. A China tem ética para o uso de dados e inteligência artificial, mas o poder decisório irrestrito é do governo
 6. A autorregulamentação falhou: cabe ao poder público a tarefa de regular a IA
9. IA e saúde
 1. Protagonismo da inteligência artificial no setor de saúde: restaurar a conexão médico-paciente
 2. Internet dos corpos: o corpo humano como plataforma tecnológica
 3. Novas fronteiras da inteligência artificial: IA emocional e chatbot terapêutico

4. Nem tudo é tecnologia: são múltiplas as formas prazerosas de escape
10. IA no combate à covid-19
 1. Sistemas de vigilância digital: combatem o coronavírus, mas ameaçam a privacidade e a liberdade
 2. O protagonismo da inteligência artificial no combate à covid-19
 3. As limitações da inteligência artificial no enfrentamento da covid-19
 4. O cenário futurista de E. M. Forster e a covid-19: a vida cotidiana transcorre mediada pela Máquina
 5. Epidemiologia digital no combate à covid-19: inovando no monitoramento de epidemias
 6. A inteligência artificial não é inteligente, é pura estatística
11. A IA e o clima
 1. Consumo de energia e emissão de CO2 dos algoritmos de inteligência artificial: como evitar uma catástrofe climática
 2. Inteligência artificial não é inteligente nem artificial
 3. O futuro é verde digital
12. IA e cultura
 1. Dá para fazer arte com inteligência artificial?
 2. Avanços e diversidade na interface arte e tecnologia: criatividade, inteligência e consciência
13. Interação humano-máquina
 1. Os algoritmos de inteligência artificial estão afetando nossa capacidade de decisão?
 2. Estamos viciados em tecnologia ou apenas em transição para o futuro?
 3. Consciência distingue a inteligência humana da inteligência de máquina?
 4. Vivências virtuais: de James Joyce ao smartphone
 5. Num ambiente complexo, nem a transparência é consensual
 6. O paradoxo da solidão na sociedade hiperconectada: os robôs inteligentes e as novas famílias
 7. Antropomorfismo na inteligência artificial: equívocos na apropriação da linguagem humana
 8. Contrariando as previsões de James Lovelock para o Novaceno
 9. Fortnite é rede social e futuro metaverso: sofisticada-se a disputa pela atenção do usuário
14. Pensando o futuro
 1. Imaginário distópico da inteligência artificial não é

- construído só pela ficção científica
2. Equipes interdisciplinares: não basta “juntar campos”, é preciso construir pontes
3. Supervalorização e demonização da inteligência artificial
4. Nem tudo é tecnologia e seus efeitos
5. A IA migrou definitivamente da academia para o mundo real: o desafio é como enfrentar esse protagonismo
6. Maior ameaça militar contemporânea: drones autônomos letais
15. Posfácio | Davi Geiger
16. Agradecimentos
17. Glossário
18. Bibliografia

Landmarks

1. Cover